PATENT ABSTRACTS OF JAPAN

(11)Publication number : **2000-242434**
(43)Date of publication of application : **08.09.2000**

(51)Int.Cl. **G06F 3/06**

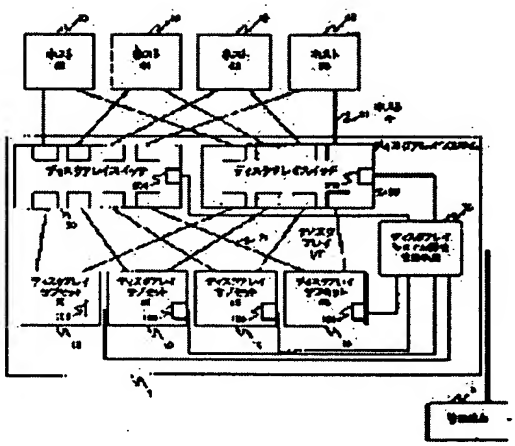
(21)Application number : **11-344260**
(22)Date of filing : **03.12.1999**

(71)Applicant : **HITACHI LTD**
(72)Inventor : **MATSUNAMI NAOTO**
OEDA TAKASHI
YAMAMOTO AKIRA
AJIMATSU YASUYUKI
SATO MASAHIKO

(30)Priority
Priority number : 10364079 Priority date : **22.12.1998** Priority country : JP

(54) STORAGE DEVICE SYSTEM

(57)Abstract:
PROBLEM TO BE SOLVED: To construct a storage device system corresponding to the scale or request of a computer system so that the extension of a storage device system and improvement in reliability in the future are easily realized.
SOLUTION: This system 1 has a plurality of subsets 10 having a storage device for holding data and a controller for controlling the storage device and switch devices 20 arranged between the subsets 10 and a host 30. Each switch device 20 has a managing table for holding management information for managing the configuration of the storage device system 1. According to the management information, address information contained in frame information outputted by the host 30 is translated and the frame information is distributed to the subsets 10.



LEGAL STATUS

[Date of request for examination] 24.12.2003

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]
[Number of appeal against examiner's decision of rejection]
[Date of requesting appeal against examiner's decision of rejection]
[Date of extinction of right]

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号
特開2000－242434
（P2000－242434A）
(43)公開日　平成12年９月８日(2000.9.8)

(51)Int.Cl.⁷	識別記号	ＦＩ	テーマコード（参考）
Ｇ０６Ｆ　3/06	３０１ ５４０	Ｇ０６Ｆ　3/06	３０１Ｇ ５４０

審査請求　未請求　請求項の数20　ＯＬ　（全 24 頁）

(21)出願番号　特願平11－344260

(22)出願日　平成11年12月３日(1999.12.3)

(31)優先権主張番号　特願平10－364079
(32)優先日　平成10年12月22日(1998.12.22)
(33)優先権主張国　日本（ＪＰ）

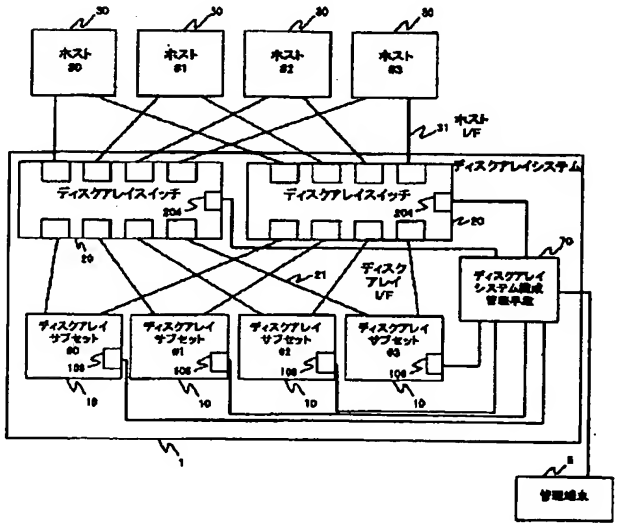
(71)出願人　000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目６番地
(72)発明者　松並　直人
神奈川県川崎市麻生区王禅寺1099番地　株式会社日立製作所システム開発研究所内
(72)発明者　大枝　高
神奈川県川崎市麻生区王禅寺1099番地　株式会社日立製作所システム開発研究所内
(74)代理人　100075096
弁理士　作田　康夫

最終頁に続く

(54)【発明の名称】　記憶装置システム

(57)【要約】
【課題】計算機システムの規模、要求などに応じた記憶装置システムを構築でき、将来における記憶装置システムの拡張、信頼性の向上を容易に実現できるようにする。
【解決手段】記憶装置システム１は、データを保持する記憶装置とそれを制御する制御装置を有する複数のサブセット１０とサブセット１０とホスト３０との間に配置されるスイッチ装置２０を有する。スイッチ装置２０は、記憶装置システム１の構成を管理する管理情報を保持する管理テーブルを有し、管理情報に従ってホスト３０が出力するフレーム情報に含まれるアドレス情報を変換してフレーム情報をサブセット１０に振り分ける。

図1



【特許請求の範囲】

【請求項１】データを保持する記憶媒体を有する記憶装置と該記憶装置を制御する制御装置とを有する複数の記憶装置サブシステムと、前記複数の記憶装置サブシステムに保持されるデータを使用する計算機に接続され、記憶装置システムの構成情報を格納した構成管理テーブルと、前記計算機から送られてくるフレームに応答して、該フレームを解析し、前記構成管理テーブルに保持された構成情報に基づいて前記フレームを変換するフレーム変換手段とを有する第１のインタフェースノードと、各々が前記記憶装置サブシステムのいずれか１つに接続された複数の第２のインタフェースノードと、前記第１のインタフェースノード及び前記複数の第２のインタフェースノードが接続され、前記第１のインタフェースノードと前記複数の第２のインタフェースノードとの間で前記フレームの転送を行う転送手段とを有することを特徴とする記憶装置システム。

【請求項２】前記第１のインタフェースノードは、前記フレームに前記第２のインタフェースノードのノードアドレス情報を付加して出力するパケット生成手段を有し、前記転送手段は、前記ノードアドレス情報に基づいて前記第１のインタフェースノードと前記複数の第２のインタフェースノードとの間で前記フレームの転送を行うことを特徴とする請求項１記載の記憶装置システム。

【請求項３】前記フレームは、転送元及び転送先を指定する識別子を保持するヘッダ部と、転送される実体データを保持するデータ実体部とを有し、前記変換手段は、前記構成情報に基づき前記ヘッダ部に保持された転送先の識別子を変換することを特徴とする請求項１記載の記憶装置システム。

【請求項４】前記フレームは、前記データ実体部に、前記計算機により認識されている第１の論理アドレス情報を含み、前記変換手段は、前記構成管理テーブルに保持された前記構成情報に基づいて、前記第１の論理アドレス情報を、該フレームの転送先となる記憶装置サブシステム内で管理される第２の論理アドレスに変換することを特徴とする請求項３記載の記憶装置システム。

【請求項５】前記記憶装置システムは、さらに、前記転送手段に接続し、オペレータから記憶装置システムの構成を定義する構成情報の入力を受け付け、該入力に応答して、各ノードの前記構成管理テーブルに前記構成情報を設定する管理プロセッサを有することを特徴とする請求項１記載の記憶装置システム。

【請求項６】前記構成情報は、前記計算機から前記複数の記憶装置サブシステムへのアクセスを制限する情報を含むことを特徴とする請求項５記載の記憶装置システム。

【請求項７】前記第１のインタフェースノードは、前記計算機から転送されてくるデータの書き込みを指示するライトコマンドフレームに応答して、該ライトコマンドフレーム及びそれに続くデータフレームについてそれらの複製を生成し、前記ライトコマンドフレーム及びそれに続くデータフレームが少なくとも２つの記憶装置サブシステムに送られるよう、各々のフレームに異なるノードアドレス情報を付加して前記転送手段に転送することを特徴とする請求項２記載の記憶装置システム。

【請求項８】前記第１のインタフェースノードは、前記計算機から転送されてくるデータのリードを指示するリードコマンドフレームに応答して、該リードコマンドフレームの複製を生成し、前記少なくとも２つの記憶装置サブシステムに前記リードコマンドフレームが送られるように、各々のリードコマンドフレームに異なるノードアドレス情報を付加して前記転送手段に転送することを特徴とする請求項７記載の記憶装置サブシステム。

【請求項９】前記第１のインタフェースノードは、前記リードコマンドフレームに応答して前記少なくとも２つの記憶装置サブシステムから転送されてくるデータフレームを受信し、その一方を選択して前記計算機に転送することを特徴とする請求項８記載の記憶装置システム。

【請求項１０】前記第１のインタフェースノードは、前記計算機から転送されてくるデータのリードを指示するリードコマンドフレームに応答して、前記少なくとも２つの記憶装置サブシステムのうち予め定められた一の記憶装置サブシステムに接続する第２のインタフェースノードのノードアドレス情報を前記リードコマンドフレームに付加して前記転送手段に転送することを特徴とする請求項７記載の記憶装置サブシステム。

【請求項１１】データを保持する記憶媒体を有する記憶装置、及び該記憶装置を制御する制御装置とを有する複数の記憶装置サブシステムと、前記記憶装置に格納されたデータを利用する計算機との間に接続されるスイッチ装置であって、前記計算機に接続され、記憶装置システムの構成情報を格納した構成管理テーブルと、前記計算機から送られてくるフレームに応答して、該フレームを解析し、前記構成管理テーブルに保持された前記構成情報に基づいて前記フレームを変換する変換手段と、各々が前記記憶装置サブシステムのいずれかに接続された複数の第２のインタフェースノードと、前記第１のインタフェースノード及び前記複数の第２のインタフェースノードが接続され、前記第１のインタフェースノードと前記複数の第２のインタフェースノードとの間で前記フレームの転送を行う転送手段とを有すること特徴とするスイッチ装置。

【請求項１２】前記第１のインタフェースノードが、前記フレームに前記第２のインタフェースノードのノードアドレス情報を付加して出力するパケット生成手段を有し、前記転送手段は、前記ノードアドレス情報に基づいて前記第１のインタフェースノードと前記複数の第２のインタフェースノードとの間で前記フレームの転送を行うことを特徴とする請求項１１記載のスイッチ装置。

【請求項１３】前記フレームは、転送元及び転送先を指定する識別子を保持するヘッダ部と、転送される実体データを保持するデータ実体部とを有し、前記変換手段は、前記構成情報に基づき前記ヘッダ部に保持された転送先の識別子を変換することを特徴とする請求項１１記載のスイッチ装置。

【請求項１４】前記フレームは、前記データ実体部に、前記計算機により認識されている前記データの格納先を示す第１の論理アドレス情報を含み、前記変換手段は、前記構成管理テーブルに保持された前記構成情報に基づいて、前記第１の論理アドレス情報を、該フレームの転送先となる記憶装置サブシステム内で管理される第２の論理アドレスに変換することを特徴とする請求項１３記載のスイッチ装置。

【請求項１５】前記スイッチ装置は、さらに、前記転送手段に接続し、オペレータから該スイッチ装置及び前記複数の記憶装置サブシステムを含んで構成される記憶装置システムの構成を定義する構成情報の入力を受け付け、該入力に応答して、各ノードの構成管理テーブルに前記構成情報を設定する管理プロセッサを有することを特徴とする請求項１１記載のスイッチ装置。

【請求項１６】前記第１のインタフェースノードは、前記計算機から転送されてくるデータの書き込みを指示するライトコマンドフレームに応答して、該ライトコマンドフレーム及びそれに続くデータフレームについてそれらの複製を生成し、前記ライトコマンドフレーム及びそれに続くデータフレームが少なくとも２つの記憶装置サブシステムに送られるよう、各々のフレームに異なるノードアドレス情報を付加して前記転送手段に転送することを特徴とする請求項１２記載のスイッチ装置。

【請求項１７】前記第１のインタフェースノードは、前記計算機から転送されてくるデータのリードを指示するリードコマンドフレームに応答して、該リードコマンドフレームの複製を生成し、前記少なくとも２つの記憶装置サブシステムに前記リードコマンドフレームが送られるように、各々のリードコマンドフレームに異なるノードアドレス情報を付加して前記転送手段に転送することを特徴とする請求項１６記載のスイッチ装置。

【請求項１８】前記第１のインタフェースノードは、前記リードコマンドフレームに応答して前記少なくとも２つの記憶装置サブシステムから転送されてくるデータフレームを受信し、その一方を選択して前記計算機に転送することを特徴とする請求項１７記載のスイッチ装置。

【請求項１９】前記第１のインタフェースノードは、前記計算機から転送されてくるデータのリードを指示するリードコマンドフレームに応答して、前記少なくとも２つの記憶装置サブシステムのうち予め定められた一の記憶装置サブシステムに接続する第２のインタフェースノードのノードアドレス情報を前記リードコマンドフレームに付加して前記転送手段に転送することを特徴とする請求項１６記載のスイッチ装置。

【請求項２０】データを保持する記憶媒体を有する記憶装置と、該記憶装置を制御する制御装置とを有する複数の記憶装置サブシステムと、前記複数の記憶装置サブシステムに保持されるデータを使用する計算機に接続された第１のインタフェースノードと、各々が前記記憶装置サブシステムのいずれか１つに接続された複数の第２のインタフェースノードと、前記第１のインタフェースノード及び前記複数の第２のインタフェースノードが接続され、前記第１のインタフェースノードと前記複数の第２のインタフェースノードとの間でフレームの転送を行う転送手段と、前記転送手段に接続し、オペレータにより入力された記憶装置システムの構成を定義する構成情報を保持する管理テーブルを備えて前記構成情報に基づいて該記憶装置システムの構成を管理する管理プロセッサとを有することを特徴とする記憶装置システム。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、複数のディスク装置を制御するディスク制御システムの実現方法に関し、特に、ディスク制御システムの高速化、低コスト化、コストパフォーマンスの向上の方法に関する。

【０００２】

【従来の技術】計算機システムに用いられる記憶装置システムとして、複数のディスク装置を制御するディスクアレイシステムがある。ディスクアレイシステムについては、例えば、“A Case for Redundant Arrays of Inexpensive Disks (RAID)”; InProc. ACM SIGMOD, June 1988（カリフォルニア大学バークレー校発行）に開示されている。ディスクアレイは、複数のディスク装置を並列に動作させることで、ディスク装置を単体で用いた記憶装置システムに比べ高速化を実現する技術である。

【０００３】複数のディスクアレイシステムを、複数のホストと相互に接続する方法として、ファイバチャネル（Fibre Channel）のFabricを使用した方法がある。この方法を適用した計算機システムの例が、日経エレクトロニクス1995.7.3（no.639）「シリアルSCSIがいよいよ市場へ」P.79 図３に示されている。ここに開示される計算機システムでは、複数のホストコンピュータ（以下では単にホストと呼ぶ）と複数のディスクアレイシステムが、それぞれ、ファイバチャネルを介してファブリック装置に接続される。ファブリック装置は、ファイバチャネルのスイッチであり、ファブリック装置に接続する任意の装置間の転送路の接続を行う。ファブリック装置はファイバチャネルのパケットである「フレーム」の転送に対し透過であり、ホストとディスクアレイシステムは、互いにファブリック装置を意識することなく２点間で通信を行う。

【０００４】

【発明が解決しようとする課題】従来のディスクアレイ

システムでは、大容量化のためディスク装置の台数を増やし、高性能化のため台数に見合った性能を有するコントローラを実現しようとすると、コントローラの内部バスの性能限界や、転送制御を行うプロセッサの性能限界が顕在化する。このような問題に対処するために、内部バスを拡張し、プロセッサ数を増加することが行われている。しかし、このような対処の仕方は、多数のバス制御によるコントローラ構成の複雑化や、プロセッサ間の共有データの排他制御等による制御ソフトの複雑化とオーバヘッドの増加を招く。このため、コストを非常に上昇させるとともに、性能は頭打ちになり、その結果、コストパフォーマンスが悪化する。また、このような装置は、大規模なシステムでは、そのコストに見合った性能が実現できるものの、規模がそれほど大きくないシステムには見合わない、拡張性が制限される、開発期間の増大と開発コストの上昇を招くといった課題がある。

【０００５】複数のディスクアレイシステムを並べファブリック装置で相互接続することによって、システム全体としての大容量化、高性能化を行うことが可能である。しかし、この方法では、ディスクアレイシステム間に関連性は全くなく、特定のディスクアレイシステムにアクセスが集中したとしてもそれを他の装置に分散することができないので、実使用上の高性能化が実現できない。また、ホストから見た論理的なディスク装置（論理ユニットと呼ぶ）の容量は、１台のディスクアレイシステムの容量に制限されるので、論理ユニットの大容量化は実現できない。

【０００６】ディスクアレイシステム全体を高信頼化しようとした際に、ホストが備えているミラーリング機能を用いて２台のディスクアレイシステムによるミラー構成を実現することができるが、ホストによるミラーリングのための制御オーバヘッドが発生し、システム性能が制限されるという課題がある。また、多数のディスクアレイシステムがシステム内に個別に存在すると、システム管理者が管理するための負荷が増加する。このため、多数の保守人員、複数台分の保守費用が必要になる等、管理コストが増加する。さらに、複数のディスクアレイシステム、ファブリック装置は、それぞれ独立した装置であるので、各種設定は、それぞれの装置毎に異なる方法で実施する必要がある。このため、管理者のトレーニングや、操作時間の増大にともない運用コストが増大する。

【０００７】本発明の目的は、これら従来技術における課題を解決し、計算機システムの規模、要求などに応じた記憶装置システムを構築でき、将来における記憶装置システムの拡張、信頼性の向上などに容易に対応することのできる記憶装置システムを実現することにある。

【０００８】

【課題を解決するための手段】本発明の記憶装置システムは、データを保持する記憶媒体を有する記憶装置と、この記憶装置を制御する制御装置とを有する複数の記憶装置サブシステム、複数の記憶装置サブシステムに保持されるデータを使用する計算機に接続された第１のインタフェースノード、各々が記憶装置サブシステムのいずれかに接続された複数の第２のインタフェースノード、及び第１のインタフェースノード及び複数の第２のインタフェースノードが接続され、第１のインタフェースノードと複数の第２のインタフェースノードとの間でフレームの転送を行う転送手段を有する。

【０００９】好ましくは、第１のインタフェースノードは、記憶装置システムの構成情報を格納した構成管理テーブルと、計算機から送られてくるフレームに応答して、該フレームを解析し、構成管理テーブルに保持された構成情報に基づいてそのフレームの転送先に関する情報変換して転送手段に転送する。

【００１０】また、フレームの転送に際して、第１のインタフェースノードは、そのフレームを受け取るべきノードのノードアドレス情報をフレームに付加する。転送手段はフレームに付加されたノードアドレス情報に従ってフレームを転送する。第２のインタフェースノードは、転送手段から受け取ったフレームからノードアドレス情報を除いてフレームを再形成し、目的の記憶装置サブシステムに転送する。

【００１１】本発明のある態様において、記憶装置システムは、転送手段に接続する管理プロセッサを有する。管理プロセッサは、オペレータからの指示に従って、構成管理テーブルに構成情報を設定する。構成情報には、計算機からのアクセスを制限する情報が含まれる。

【００１２】

【発明の実施の形態】［第１実施形態］図１は、本発明が適用されたディスクアレイシステムを用いたコンピュータシステムの一実施形態における構成図である。

【００１３】1はディスクアレイシステム、３０はディスクアレイシステムが接続されるホストコンピュータ（ホスト）である。ディスクアレイシステム1は、ディスクアレイサブセット１０、ディスクアレイスイッチ２０、ディスクアレイシステム全体の設定管理を行うディスクアレイシステム構成管理手段７０、ディスクアレイスイッチ２０とディスクアレイシステム構成管理手段７０との間、およびディスクアレイサブセット１０ディスクアレイシステム構成管理手段７０との間の通信インタフェース（通信Ｉ／Ｆ）８０を有する。ホスト３０とディスクアレイシステム1とは、ホストインタフェース（ホストＩ／Ｆ）３１で接続されており、ホストＩ／Ｆ３１はディスクアレイシステム1のディスクアレイスイッチ２０に接続する。ディスクアレイシステム1の内部において、ディスクアレイスイッチ２０とディスクアレイサブセット１０は、ディスクアレイインタフェース（ディスクアレイＩ／Ｆ２１）で接続される。

【００１４】ホスト３０、ディスクアレイサブセット１

0は、図では、各々４台示されているが、この台数に関しては制限はなく任意である。ホスト３０とディスクアレイサブセット１０の台数が異なっても構わない。また、ディスクアレイスイッチ２０は、本実施形態では図示の通り二重化されている。各ホスト３０および各ディスクアレイサブセット１０は、それぞれ別々のホストＩ／Ｆ３１、ディスクアレイＩ／Ｆ２１で二重化されたディスクアレイスイッチ２０の双方に接続されている。これは、一方のディスクアレイスイッチ２０、ホストＩ／Ｆ３１、あるいはディスクアレイＩ／Ｆ２１が故障しても他方を使用することでホスト３０からディスクアレイシステム1へのアクセスを可能とし、高い可用性を実現するためである。しかし、このような二重化は必ずしも必須ではなく、システムに要求される信頼性レベルに応じて選択可能である。

【００１５】図2は、ディスクアレイサブセット１０の一構成例を示す構成図である。１０１は上位システム（ホスト１０）からのコマンドを解釈してキャッシュヒットミス判定を実施し、上位システムとキャッシュ間のデータ転送を制御する上位アダプタ、１０２はディスクデータアクセス高速化のためのキャッシュ、および、マルチプロセッサ間の共有データを格納する共有メモリ（以下キャッシュ・共有メモリと呼ぶ）、１０４はディスクアレイサブセット10内に格納される複数のディスクユニットである。１０３はディスクユニット１０４を制御し、ディスクユニット１０４とキャッシュ間のデータ転送を制御する下位アダプタである。１０６はディスクアレイサブセット構成管理手段であり、ディスクアレイシステム1全体を管理するディスクアレイシステム構成管理手段７０と通信Ｉ／Ｆ８０を介して通信し、構成パラメータの設定や、障害情報の通報等の管理を行う。

【００１６】上位アダプタ１０１、キャッシュ・共有メモリ１０２、下位アダプタ１０３はそれぞれ二重化されている。この理由は上記ディスクアレイスイッチ２０の二重化と同様、高可用性を実現するためであり必須ではない。また、各ディスクユニット１０４は、二重化された下位アダプタ１０３のいずれからも制御可能である。本実施形態では、低コスト化の観点から同一のメモリ手段をキャッシュと共有メモリに共用しているが、これらは勿論分離することも可能である。

【００１７】上位アダプタ１０１は、上位アダプタ１０１の制御を実行する上位ＭＰＵ１０１０、上位システム、すなわちディスクアレイスイッチ２０との接続Ｉ／Ｆであるディスクアレイ Ｉ／Ｆ２１を制御するディスクアレイＩ／Ｆコントローラ１０１１、キャッシュ・共有メモリ１０２と上位ＭＰＵ１０１０とディスクアレイＩ／Ｆコントローラ１０１１との間の通信、データ転送を行う上位バス１０１２を含む。

【００１８】図では各上位アダプタ１０１毎に1台のディスクアレイＩ／Ｆコントローラ１０１１が示されているが、１つの上位アダプタに対し、複数のディスクアレイＩ／Ｆコントローラ１０１１を設けてもよい。

【００１９】下位アダプタ１０３は、下位アダプタ１０３の制御を実行する下位ＭＰＵ１０３０、ディスク１０４とのインタフェースであるディスクＩ／Ｆを制御するディスクＩ／Ｆコントローラ１０３１、キャッシュ・共有メモリ１０２と下位ＭＰＵ１０３０とディスクＩ／Ｆコントローラ１０３１との間の通信、データ転送を行う下位バス１０３２を含む。

【００２０】図では各下位アダプタ１０３毎に4台のディスクＩ／Ｆコントローラ１０３１が示されているが、その数は任意であり、ディスクアレイの構成や、接続するディスク台数に応じて変更可能である。

【００２１】図3は、ディスクアレイスイッチ２０の一構成例を示す構成図である。２００はディスクアレイスイッチ全体の制御および管理を行うプロセッサである管理プロセッサ（ＭＰ）、２０１はｎ×ｎの相互スイッチ経路を構成するクロスバスイッチ、２０２はディスクアレイＩ／Ｆ２１毎に設けられるディスクアレイＩ／Ｆノード、２０３はホストＩ／Ｆ３１毎に設けられるホストＩ／Ｆノード、２０４はディスクアレイシステム構成管理手段７０との間の通信を行う通信コントローラである。２０２０はディスクアレイＩ／Ｆノード２０２とクロスバスイッチ２０１を接続するパス、２０３０はホストＩ／Ｆノード２０３とクロスバスイッチ２０１を接続するパス、２０４０は他のディスクアレイスイッチ２０と接続し、クラスタを構成するためのクラスタ間Ｉ／Ｆ、２０５０はＭＰ２００とクロスバスイッチ２０１を接続するためのパスである。

【００２２】図4はクロスバスイッチ２０１の構造を示す構成図である。２０１０はクロスバスイッチ２０１に接続するパス２０２０、２０３０、２０５０、およびクラスタ間Ｉ／Ｆ２０４０を接続するポートであるスイッチングポート（ＳＷＰ）である。ＳＷＰ２０１０はすべて同一の構造を有し、あるＳＷＰから他のＳＷＰへの転送経路のスイッチング制御を行う。図では１つのＳＷＰについてのみ転送経路を示しているが、すべてのＳＷＰ間で同様の転送経路が存在する。

【００２３】図5は、ホストＩ／Ｆノード２０３の一構成例を示す構成図である。本実施形態では、具体的に説明をするためにホストＩ／Ｆ３１とディスクアレイＩ／Ｆ２１の両方にファイバチャネルを使用するものと仮定する。もちろんホストＩ／Ｆ３１とディスクアレイＩ／Ｆ２１として、ファイバチャネル以外のインタフェースを適用することも可能である。ホストＩ／Ｆノード２０３とディスクアレイＩ／Ｆノード２０２の両方に同一のインタフェースを使用することで、両者を同一構造にできる。本実施形態においては、ディスクアレイＩ／Ｆノード２０２も図に示すホストＩ／Ｆノード２０３と同様に構成される。以下では、ホストＩ／Ｆノード２０３を

例に説明を行う。

【0024】2021は受信したファイバチャネルフレーム（以下単にフレームと呼ぶ）をどのノードに転送するかを検索する検索プロセッサ（SP）、2022はホスト30（ディスクアレイI／Fノード202の場合は、ディスクアレイサブセット10）との間でフレームを送受信するインタフェースコントローラ（IC）、2022はIC2023が受信したフレームに対しSP2021が検索した結果に基づいて変換を施すスイッチングコントローラ（SC）、2024はSC2021が変換したフレームを他のノードに転送するためにクロスバスイッチ201を通過できる形式にパケット化するパケット生成部（SPG）、2025は受信したフレームを一時的に格納するフレームバッファ（FB）、2026は一つのホストからのディスクアレイアクセス要求コマンド（以下単にコマンドと呼ぶ）に対応した複数のフレーム列であるエクスチェンジ（Exchange）を識別するためのエクスチェンジ番号を管理するエクスチェンジテーブル（ET）、2027は複数のディスクアレイサブセット10の構成情報を格納するディスクアレイ構成管理テーブル（DCT）である。

【0025】ディスクアレイスイッチ20の各構成部は、すべてハードウェアロジックで構成されることが性能上望ましい。しかし、求められる性能を満足できるならば、汎用プロセッサを用いたプログラム制御によりSP2021やSC2022の機能を実現することも可能である。

【0026】各ディスクアレイサブセット10は、各々が有するディスクユニット104を1または複数の論理的なディスクユニットとして管理している。この論理的なディスクユニットを論理ユニット（LU）と呼ぶ。LUは、物理的なディスクユニット104と1対1で対応する必要はなく、1台のディスクユニット104に複数のLUが構成され、あるいは、複数のディスクユニット104で1つのLUが構成されても構わない。

【0027】ディスクアレイサブセット10の外部から見た場合、1つのLUは、1台のディスク装置として認識される。本実施形態では、ディスクアレイスイッチ20によりさらに論理的なLUが構成され、ホスト30は、このLUに対してアクセスするように動作する。本明細書では、1つのLUでホスト30から認識される1つのLUが構成される場合、ホスト30により認識されるLUを独立LU（ILU）、複数のLUでホスト30から認識される1つのLUが構成される場合、ホスト30により認識されるLUを統合LU（CLU）と呼ぶ。

【0028】図12に、4つのディスクアレイサブセットのLUで1つの統合LUが構成される場合における各階層間でのアドレス空間の対応関係を示す。図において、1000は、一例として、ホスト“#2”からみたディスクアレイシステム1の1つの統合LUにおけるアドレス空間、1100は、ディスクアレイサブセット10のLUのアドレス空間、1200はディスクユニット104（ここでは、ディスクアレイサブセット“#0”についてのみ図示されている）のアドレス空間を示している。

【0029】各ディスクアレイサブセット10のLUは、ここでは、4台のディスクユニット104によりRAID5（Redundant Arrays of Inexpensive Disks Level 5）型ディスクアレイとして構成されるものとする。各ディスクアレイサブセット10は、それぞれn0、n1、n2、n3の容量を有するLUを持つ。ディスクアレイスイッチ20は、これら4つのLUの持つアドレス空間を（n0+n1+n2+n3）の容量を有するアドレス空間に統合し、ホスト30から認識される統合LUを実現する。

【0030】本実施形態では、例えば、ホスト#2が領域A1001をアクセスする場合、領域A1001を指定したアクセス要求は、ディスクアレイスイッチ20によりディスクアレイサブセット#0のLUの領域A′1101をアクセスするための要求に変換されてディスクアレイサブセット#0に転送される。ディスクアレイサブセット#0は、領域A′1101をさらに、ディスクユニット104上の領域A″1201にマッピングしてアクセスを行う。アドレス空間1000とアドレス空間1100との間のマッピングは、ディスクアレイスイッチ20が有するDCT207に保持された構成情報に基づき行われる。この処理の詳細については後述する。なお、ディスクアレイサブセット内におけるマッピングについては、既によく知られた技術であり、本明細書では詳細な説明については省略する。

【0031】本実施形態において、DCT207は、システム構成テーブルとサブセット構成テーブルを含む。図6は、システム構成テーブルの構成を、図7は、サブセット構成テーブルの構成を示す。

【0032】図7に示すように、システム構成テーブル20270は、ホストLUの構成を示す情報を保持するホストLU構成テーブル20271、及びディスクアレイスイッチ20のディスクアレイI／Fノード202とディスクアレイサブセット10との接続関係を示すディスクアレイI／Fノード構成テーブル20272を有する。

【0033】ホストLU構成テーブル20271は、ホスト30からみたLUごとに、そのLUを識別する番号であるHost-LU No.、LUの属性を示すLU Type、CLU Class、及びCLU Stripe Size、ホストLUの状態を示す情報であるCondition、ホストLUを構成するディスクアレイサブセット10のLUに関する情報であるLU情報（LU Info.）を有する。

【0034】LU Typeは、このホストLUがCLUであるか、ILUであるかといったLUの種類を示す情報で

ある。CLU Classは、LU TypeによりこのホストＬＵがＣＬＵであることが示される場合に、そのクラスが“Joined”、“mirrored”、及び“Striped”のいずれであるかを示す情報である。“Joined”は、図１１により説明したように、いくつかのＬＵを連結して１つの大きな記憶空間を持つＣＬＵが構成されていることを示す。“Mirrored”は、第６実施形態として後述するように、２つのＬＵにより二重化されたＬＵであることを示す。“Striped”は、第７実施形態として後述するように、複数のＬＵで構成され、データがこれら複数のＬＵに分散して格納されたＬＵであることを示す。CLU Stripe Sizeは、CLU Classにより「Striped」であることが示される場合に、ストライピングサイズ（データの分散の単位となるブロックのサイズ）を示す。

【００３５】Conditionにより示される状態には、“Normal”、“Warning”、“Fault”、及び“Not Defined”の４種類がある。“Normal”はこのホストＬＵが正常な状態であることを示す。“Warning”は、このホストＬＵを構成するＬＵに対応するいずれかのディスクユニットに障害が発生している等の理由により縮退運転が行われていることを示す。“Fault”は、ディスクアレイサブセット１０の故障などによりこのホストＬＵを運転することができないことを示す。“Not Defined”は、対応するHost-LU No.のホストＬＵが定義されていないことを示す。

【００３６】LU Infoは、このホストＬＵを構成するＬＵについて、そのＬＵが属するディスクアレイサブセット１０を特定する情報、ディスクアレイサブセット内でのＬＵＮ、及びそのサイズを示す情報を含む。ホストＬＵがＩＬＵの場合には、唯一のＬＵに関する情報が登録される。ホストＬＵがＣＬＵの場合には、それを構成する全てのＬＵについて、それぞれのＬＵに関する情報が登録される。例えば、図において、Host-LU No.が“０”であるHost-LUは、ディスクアレイサブセット“＃０”のＬＵＮ“０”、ディスクアレイサブセット“＃１”のＬＵＮ“０”、ディスクアレイサブセット“＃２”のＬＵＮ“０”、ディスクアレイサブセット“＃３”のＬＵＮ“０”の4つのＬＵから構成されるＣＬＵであり、そのＣＬＵクラスが“Joined”であるＣＬＵであることが分かる。

【００３７】ディスクアレイＩ／Ｆノード構成テーブル20272は、ディスクアレイＩ／Ｆ２１が接続するディスクアレイサブセット１０のポートごとに、どのディスクアレイスイッチ２０のディスクアレイＩ／Ｆノード２０２が接続されるかを示す情報を保持する。

【００３８】具体的には、ディスクアレイサブセット１０を特定するSubset No.、ポートを特定するSubset Port No.、そのポートに接続するディスクアレイスイッチ２０を特定するSwitch No.、及びそのディスクアレイスイッチ２０のディスクアレイＩ／Ｆノード２０２を特定するI/F Node No.を有する。ディスクアレイサブセット１０が複数のポートを備えている場合には、そのポート毎に情報が設定される。

【００３９】サブセット構成テーブルは、図７に示すように、各ディスクアレイサブセット１０に対応する複数のテーブル202720～202723を有する。各テーブルは、ディスクアレイサブセット１０内で構築されたＲＡＩＤグループの構成を示す情報を保持するＲＡＩＤグループ構成テーブル202730と、ディスクアレイサブセット１０内に構築されたＬＵの構成を示す情報を保持するＬＵ構成テーブル202740を含む。

【００４０】ＲＡＩＤグループ構成テーブル202730は、ＲＡＩＤグループに付加された番号を示すGroup No.、そのＲＡＩＤグループのレベルを示すLevel、そのＲＡＩＤグループを構成するディスクの数を示す情報であるDisks、そのＲＡＩＤグループがＲＡＩＤレベル０，５等のストライピングされた構成の場合、そのストライプサイズを示すStripe Sizeを情報として含む。例えば、図に示されるテーブルにおいて、ＲＡＩＤグループ“０”は、４台のディスクユニットにより構成されたＲＡＩＤグループであり、ＲＡＩＤレベルが５、ストライプサイズがＳ０である。

【００４１】ＬＵ構成テーブル202740は、ＬＵに付加された番号（ＬＵＮ）を示すLU No.、このＬＵがどのＲＡＩＤグループに構成されているのかを示すRAID Group、ＬＵの状態を示すCondition、このＬＵのサイズ（容量）を示すSize、このＬＵがディスクアレイサブセット１０のどのポートからアクセス可能なのかを示すPort、及びその代替となるポートを示すAlt. Portを情報として含む。Conditionで示される状態は、ホストＬＵについてのConditionと同様、“Normal”、“Warning”、“Fault”、“Not Defined”の４種類がある。Alt. Portに設定された情報により特定されるポートは、Portに設定された情報で特定されるポートに障害が発生したときに用いられるが、単に複数のポートから同一のＬＵをアクセスするために用いることもできる。

【００４２】図８は、ファイバチャネルにおけるフレームの構成図である。ファイバチャネルのフレーム４０は、フレームの先頭を示すＳＯＦ（Start Of Frame）４００、フレームヘッダ４０１、転送の実態データを格納する部位であるフレームペイロード４０２、３２ビットのエラー検出コードであるＣＲＣ（Cyclic RedundancyCheck）４０３、フレームの最後尾を示すＥＯＦ（End Of Frame）４０４を含む。フレームヘッダ４０１は、図９に示すような構造になっており、フレーム転送元のＩＤ（S_ID）、フレーム転送先のＩＤ（D_ID）、エクスチェンジの起動元、応答先が指定するそれぞれのエクスチェンジＩＤ（OX_ID、RX_ID）、エクスチェンジ中のフレームグループを指定するシーケンスのＩＤ（SEQ_ID）等が格納されている。

【００４３】本実施形態では、ホスト３０により発行されるフレームには、S_IDとしてホスト３０に割り当てられたＩＤが、また、D_IDとしてディスクアレイスイッチ２０のポートに割り当てられたＩＤが使用される。一つのホストコマンドに対し、１ペアのエクスチェンジＩＤ（OX_ID、RX_ID）が割り当てられる。複数のデータフレームを同一のエクスチェンジに対し発行する必要があるときは、その全データフレームに対して同一のSEQ_IDが割り当てられ、おのおのはシーケンスカウント（SEQ_CNT）で識別される。フレームペイロード４０２の最大長は２１１０バイトであり、フレーム種毎に格納される内容が異なる。例えば、後述するFCP_CMDフレームの場合、図１０に示すように、ＳＣＳＩのLogical Unit Number（ＬＵＮ）、Command Description Block（ＣＤＢ）等が格納される。ＣＤＢは、ディスク（ディスクアレイ）アクセスに必要なコマンドバイト、転送開始論理アドレス（ＬＢＡ）、転送長（ＬＥＮ）を含む。

【００４４】以下、本実施形態のディスクアレイシステムの動作を説明する。

【００４５】ディスクアレイシステムを使用するのに先立ち、ディスクアレイスイッチ２０に対して、ディスクアレイサブセット１０の構成情報を設定する必要がある。システム管理者は、管理端末５からディスクアレイシステム構成手段７０を介して、すべてのディスクアレイサブセット１０およびディスクアレイスイッチ２０の構成設定情報を獲得する。管理者は、管理端末５から所望のシステム構成になるよう論理ユニットの構成設定、RAIDレベルの設定、障害発生時の交代パスの設定等、各種設定に必要な設定情報を入力する。ディスクアレイシステム構成管理手段７０は、その設定情報を受け、各ディスクアレイサブセット１０およびディスクアレイスイッチ２０に設定情報を転送する。なお、管理端末５における設定情報の入力については第５実施形態にて別途説明する。

【００４６】ディスクアレイスイッチ２０では、通信コントローラ２０４が設定情報を獲得し、ＭＰ２００により各ディスクアレイサブセット１０のアドレス空間情報等の構成情報が設定される。ＭＰ２００は、クロスバスイッチ２０１経由で各ホストＩ／Ｆノード２０３およびディスクアレイＩ／Ｆノード２０２に、ディスクアレイサブセット１０の構成情報を配信する。

【００４７】各ノード２０３、および２０２はこの情報を受信すると、ＳＰ２０２１により構成情報をＤＣＴ２０２７に格納する。ディスクアレイサブセット１０では、ディスクアレイサブセット構成管理手段１０６が、設定情報を獲得し、共有メモリ１０２に格納する。各上位ＭＰＵ１０１０および下位ＭＰＵ１０３０は、共有メモリ１０２上の設定情報を参照し、各々の構成管理を実施する。

【００４８】以下では、ホスト“＃２”がディスクアレイシステム1に対し、リードコマンドを発行した場合の動作を説明する。図１１に、ホストからのリード動作時にファイバチャネルを通して転送されるフレームのシーケンスを示す模式図を、図１３にこのときのディスクアレイスイッチのホストＩ／Ｆノード２０３における動作のフローチャートを示す。

【００４９】なお、以下の説明では、ホスト“＃２”が、図１２における記憶領域Ａ１００１をアクセスすることを仮定する。記憶領域Ａ１００１に対応する実際の記憶領域A″は、ディスクアレイサブセット“＃０”のＬＵＮ＝０のＬＵを構成するディスクユニット＃２のアドレス空間内に存在するものとする。また、アドレス空間１０００を構成するＬＵを定義しているホストＬＵ構成テーブル20271のLU Typeには「ＣＬＵ」が、CLU Classには「Joined」が設定されているものとする。

【００５０】データのリード時、ホスト３０は、リードコマンドを格納したコマンドフレーム「FCP_CMD」をディスクアレイスイッチ２０に発行する（図１１矢印（ａ））。ディスクアレイスイッチ２０のホストＩ／Ｆノード“＃２”は、ＩＣ２０２３によりホストＩ／Ｆ３１経由でコマンドフレーム「FCP_CMD」を受信する（ステップ20001）。ＩＣ２０２３は、ＳＣ２０２２にコマンドフレームを転送する。ＳＣ２０２２は、受け取ったコマンドフレームを一旦ＦＢ２０２５に格納する。この際、ＳＣ２０２２は、コマンドフレームのＣＲＣを計算し、受信情報が正しいことを検査する。ＣＲＣの検査に誤りがあれば、ＳＣ２０２２は、その旨をＩＣ２０２３に通知する。ＩＣ２０２３は、誤りの通知をＳＣ２０２２から受けると、ホストＩ／Ｆ３１を介してホスト３０にＣＲＣエラーを報告する。（ステップ20002）。

【００５１】ＣＲＣが正しい場合、ＳＣ２０２２は、ＦＢ２０２５に保持したフレームをリードし、それがコマンドフレームであることを認識してフレームヘッダ４０１を解析する（ステップ20003）。そして、ＳＣ２０２２は、ＳＰ２０２１に指示し、S_ID、D_ID、OX_ID等のエクスチェンジ情報をＥＴ２０２６に登録する（ステップ20004）。

【００５２】次に、ＳＣ２０２２は、フレームペイロード４０２を解析し、ホスト３０により指定されたＬＵＮおよびＣＤＢを取得する（ステップ20005）。ＳＰ２０２１は、ＳＣ２０２２の指示により、ＤＣＴ２０２７を検索し、ディスクアレイサブセット１０の構成情報を得る。具体的には、ＳＰ２０２１は、ホストＬＵ構成テーブル20271を検索し、受信したフレームペイロード４０２に格納されたＬＵＮと一致するHost-LU No.を有する情報を見つける。ＳＰ２０２１は、LU Type、CLU Classに設定された情報からホストＬＵの構成を認識し、LU Info.に保持されている情報に基づきアクセスすべきディスクサブセット１０とその中のＬＵのＬＵＮ、及びこのＬＵ内でのＬＢＡを判別する。次に、ＳＰ２０２１は、

サブセット構成テーブル202720のＬＵ構成テーブル202740を参照し、目的のディスクアレイサブセット１０の接続ポートを確認し、ディスクアレイＩ／Ｆノード構成テーブル20272からそのポートに接続するディスクアレイＩ／Ｆノード２０２のノードNo.を得る。ＳＰ２０２１は、このようにして得たディスクアレイサブセット１０を識別する番号、ＬＵＮ、ＬＢＡ等の変換情報をＳＣ２０２２に報告する。（ステップ20006）。

【００５３】次に、ＳＣ２０２２は、獲得した変換情報を使用しフレームペイロード４０２のＬＵＮとＣＤＢのなかのＬＢＡを変換する。また、フレームヘッダ４０１のD_IDを対応するディスクアレイサブセット１０のホストＩ／Ｆコントローラ１０１１のD_IDに変換する。なお、この時点ではS_IDは書き換えない（ステップ２００７）。

【００５４】ＳＣ２０２２は、変換後のコマンドフレームと、対象ディスクアレイサブセット１０に接続するディスクアレイＩ／Ｆノード番号を、ＳＰＧ２０２４に転送する。ＳＰＧ２０２４は、受け取った変換後のコマンドフレームに対し、図１４に示すような簡単な拡張ヘッダ６０１を付加したパケットを生成する。このパケットをスイッチングパケット（Ｓ　Ｐａｃｋｅｔ）６０と呼ぶ。S Packet６０の拡張ヘッダ６０１には、転送元（自ノード）番号、転送先ノード番号、及び転送長が付加含まれる。ＳＰＧ２０２４は、生成したS Packet６０をクロスバスイッチ２０１に送信する（ステップ20008）。

【００５５】クロスバスイッチ２０１は、ホストＩ／Ｆノード“＃２”と接続するＳＷＰ２０１０によりS Packet６０を受信する。ＳＷＰ２０１０は、S Packet６０の拡張ヘッダ６０１を参照し、転送先のノードが接続するＳＷＰへのスイッチ制御を行って経路を確立し、S Packet６０を転送先のディスクアレイＩ／Ｆノード２０２（ここでは、ディスクアレイＩ／Ｆノード“＃０”）に転送する。ＳＷＰ２０１０は、経路の確立をS Packet６０の受信の度に実施し、S Packet６０の転送が終了したら、その経路を解放する。ディスクアレイＩ／Ｆノード“＃０”では、ＳＰＧ２０２４がS Packet６０を受信し、拡張ヘッダ６０１を外してコマンドフレームの部分をＳＣ２０２２に渡す。

【００５６】ＳＣ２０２２は、受け取ったコマンドフレームのフレームヘッダのS_IDに自分のＩＤを書き込む。次にＳＣ２０２２は、ＳＰ２０２１に対し、コマンドフレームのS_ID、D_ID、OX_ID等のエクスチェンジ情報、及びフレーム転送元ホストＩ／Ｆノード番号をＥＴ２０２６に登録するよう指示し、ＩＣ２０２３にコマンドフレームを転送する。ＩＣ２０２３は、フレームヘッダ４０１の情報に従い、接続するディスクアレイサブセット１０（ここでは、ディスクアレイサブセット“＃０”）にコマンドフレームを転送する（図１１矢印（ｂ））。

【００５７】ディスクアレイサブセット“＃０”は、変換後のコマンドフレーム「FCP_CMD」をディスクアレイＩ／Ｆコントローラ１０１１で受信する。上位ＭＰＵ１０１０は、コマンドフレームのフレームペイロード４０２に格納されたＬＵＮとＣＤＢを取得し、指定された論理ユニットのＬＢＡからＬＥＮ長のデータをリードするコマンドであると認識する。

【００５８】上位ＭＰＵ１０１０は、共有メモリ１０２に格納されたキャッシュ管理情報を参照し、キャッシュヒットミス／ヒット判定を行う。ヒットすればキャッシュ１０２からデータ転送を実施する。ミスの場合、ディスクユニットからデータをリードする必要があるので、ＲＡＩＤ５の構成に基づくアドレス変換を実施し、キャッシュ空間を確保する。そして、ディスクユニット2からのリード処理に必要な処理情報を生成し、下位ＭＰＵ１０３０に処理を引き継ぐべく、共有メモリ１０２に処理情報を格納する。

【００５９】下位ＭＰＵ１０３０は、共有メモリ１０２に処理情報が格納されたことを契機に処理を開始する。下位ＭＰＵ１０３０は、適切なディスクＩ／Ｆコントローラ１０３１を特定し、ディスクユニット２へのリードコマンドを生成して、ディスクＩ／Ｆコントローラ１０３１にコマンドを発行する。ディスクＩ／Ｆコントローラ１０３１は、ディスクユニット2からリードしたデータをキャッシュ１０２の指定されたアドレスに格納して下位ＭＰＵ１０３０に終了報告を通知する。下位ＭＰＵ１０３０は、処理が正しく終了したことを上位ＭＰＵ１０１０に通知すべく共有メモリ１０２に処理終了情報を格納する。

【００６０】上位ＭＰＵ１０１０は、共有メモリ１０２に処理終了情報が格納されたことを契機に処理を再開し、ディスクアレイＩ／Ｆコントローラ１０１１にリードデータ準備完了を通知する。ディスクアレイＩ／Ｆコントローラ１０１１は、ディスクアレイスイッチ２０の当該ディスクアレイＩ／Ｆノード“＃０”に対し、ファイバチャネルにおけるデータ転送準備完了フレームである「FCP_XFER_RDY」を発行する（図１１矢印（ｃ））。

【００６１】ディスクアレイＩ／Ｆノード“＃０”では、データ転送準備完了フレーム「FCP_XFER_RDY」を受信すると、ＳＣ２０２２が、ディスクアレイサブセット２０から受信した応答先エクスチェンジＩＤ（RX_ID）を獲得し、S_ID、D_ID、OX_IDを指定して、ＳＰ２０２１に指示しＥＴ２０２６の当該エクスチェンジ情報にRX_IDを登録する。ＳＣ２０２２は、データ転送準備完了フレームの転送先（コマンドフレームの転送元）のホストＩ／Ｆノード番号を獲得する。ＳＣ２０２２は、このフレームのS_IDを無効化し、ＳＰＧ２０２４に転送する。ＳＰＧ２０２４は、先に述べたようにしてS Packetを生成し、クロスバスイッチ２０１経由で対象ホストＩ／Ｆノード“＃２”に転送する。

【００６２】ホストＩ／Ｆノード“＃２”では、ＳＰＧ

２０２４がデータ転送準備完了フレームのS Packetを受信すると、S Packetの拡張ヘッダを外し「FCP_XFER_RDY」を再生してＳＣ２０２２に渡す（ステップ20011）。ＳＣ２０２２は、ＳＰ２０２１に指示しＥＴ２０２６をサーチして該当するエクスチェンジを特定する（ステップ20012）。

【００６３】次に、ＳＣ２０２２は、フレームが「FCP_XFER_RDY」であるかどうか調べ（ステップ20013）、「FCP_XFER_EDY」であれば、ＥＴ２０２６の応答先エクスチェンジＩＤ（RX_ID）の更新をＳＰ２０２１に指示する。応答先エクスチェンジＩＤとしては、このフレームに付加されていた値が使用される（ステップ20014）。そして、ＳＣ２０２２は、フレームヘッダ４０１のS_ID、D_IDをホストＩ／Ｆノード２０３のＩＤとホスト３０のＩＤを用いた適切な値に変換する（ステップ20015）。これらの処理によりフレームヘッダ４０１は、ホスト“＃２”に対するフレームに変換される。ＩＣ２０２３は、ホスト“＃２”に対し、このデータ転送準備完了フレーム「FCP_XFER_RDY」を発行する（図１１の矢印（ｄ）：ステップ20016）。

【００６４】ディスクアレイサブセット“＃０”のディスクアレイＩ／Ｆコントローラ１０１１は、データ転送を行うため、データフレーム「FCP_DATA」を生成し、ディスクアレイスイッチ２０に転送する（図１１矢印（ｅ））。フレームペイロードの転送長には制限があるため、１フレームで転送できる最大のデータ長は２ＫＢである。データ長がこれを越える場合は、必要数だけデータフレームを生成し発行する。すべてのデータフレームには同一のSEQ_IDが割り当てられる。データフレームの発行は、同一のSEQ_IDに対し複数のフレームが生成されることを除き（すなわちSEQ_CNTが変化する）、データ転送準備完了フレームの場合と同様である。

【００６５】ディスクアレイスイッチ２０は、データ転送準備完了フレームの処理と同様に、データフレーム「FCP_DATA」のフレームヘッダ４０１の変換を実施する。ただし、データフレームの転送の場合、RX_IDが既に確立されているので、データ転送準備完了フレームの処理におけるステップ20014の処理はスキップされる。フレームヘッダ４０１の変換後、ディスクアレイスイッチ２０は、ホスト“＃２”にデータフレームを転送する（図１１矢印（ｆ））。

【００６６】次に、ディスクアレイサブセット“＃０”のディスクアレイＩ／Ｆコントローラ１０１１は、終了ステータス転送を行うため、ステータスフレーム「FCP_RSP」を生成し、ディスクアレイスイッチ２０に対し発行する（図１１矢印（ｇ））。ディスクアレイスイッチ２０では、データ転送準備完了フレームの処理と同様に、ＳＰＧ２０２４がS Packetから拡張ヘッダを外し「FCP_RSP」ステータスフレームを再現し（ステップ20021）、ＳＰ２０２１によりＥＴ２０２６を検索しエクスチェンジ情報を獲得する（ステップ20022）。ＳＣ２０２２は、その情報に基づきフレームを変換する（ステップ20023）。変換されたフレームは、ＩＣ２０２３によりホスト“＃２”に転送される（図１１矢印（ｈ）：ステップ20024）。最後にＳＰ２０２１は、ＥＴ２０２６からエクスチェンジ情報を削除する（ステップ20025）。

【００６７】以上のようにしてディスクアレイからのリード処理が行われる。ディスクアレイシステム１に対するライト処理についてもデータフレームの転送方向が逆転するのみで、上述したリード処理と同様の処理が行われる。

【００６８】図3に示したように、ディスクアレイスイッチ２０は、クロスバスイッチ２０１にクラスタ間Ｉ／Ｆ２０４０を備えている。図１に示したシステム構成では、クラスタ間Ｉ／Ｆ２０４０は使用されていない。本実施形態のディスクアレイスイッチ２０は、クラスタ間Ｉ／Ｆ２０４０を利用して図１５に示すように、他のディスクアレイスイッチと相互に接続されることができる。

【００６９】本実施形態におけるディスクアレイスイッチ２０単独では、ホスト３０とディスクアレイサブセット１０を合計８台までしか接続できないが、クラスタ間Ｉ／Ｆ２０４０を利用して複数のディスクアレイスイッチを相互接続し、接続できるホスト１０とディスクアレイの数を増やすことができる。例えば、図１５に示すシステムでは、４台のディスクアレイスイッチ２０を使ってホスト３０とディスクアレイサブセット１０を合計３２台まで接続でき、これらの間で相互にデータ転送が可能になる。

【００７０】このように、本実施形態では、ディスク容量や性能の必要性に合わせて、ディスクアレイサブセットやホストの接続台数を増加していくことができる。また、必要な転送帯域分のホストＩ／Ｆを用いてホストーディスクアレイシステム間を接続することができるので、容量、性能、接続台数の拡張性を大幅に向上させることができる。

【００７１】以上説明した実施形態によれば、１台のディスクアレイサブセットの性能が、内部のＭＰＵや内部バスで制限されたとしても、複数のディスクアレイサブセットを用いて、ディスクアレイスイッチによりホストとディスクアレイサブセット間を相互接続することができる。これにより、ディスクアレイシステムトータルとして高い性能を実現することができる。ディスクアレイサブセットの性能が比較的低いものであっても、複数のディスクアレイサブセットを用いることで高性能化を実現できる。したがって、低コストのディスクアレイサブセットをコンピュータシステムの規模に合わせて必要な台数だけ接続することができ、規模に応じた適切なコストでディスクアレイシステムを構築することが可能とな

る。

【００７２】また、ディスク容量の増大や性能の向上が必要になったときは、ディスクアレイサブセットを必要なだけ追加すればよい。さらに、複数のディスクアレイスイッチを用いて任意の数のホスト及びディスクアレイサブセットを接続できるので、容量、性能、接続台数のいずれをも大幅に向上させることができ、高い拡張性を有するシステムが実現できる。

【００７３】さらにまた、本実施形態によれば、ディスクアレイサブセットとして、従来のディスクアレイシステムそのものの縮小機を用いることができるので、既に開発した大規模な制御ソフトウェア資産をそのまま利用でき、開発コストの低減と開発期間の短縮を実現することができる。

【００７４】［第２実施形態］図１６は、本発明の第２の実施形態におけるコンピュータシステムの構成図である。本実施形態は、ディスクアレイスイッチのホストＩ／Ｆノードにおいて、フレームヘッダ４０１のみを変換し、フレームペイロード４０２は操作しない点、及び、ディスクアレイスイッチ、ホストＩ／Ｆ、ディスクアレイＩ／Ｆが二重化されていない点で第１実施形態と構成上相違する。したがって、各部の構成は、第１実施形態と大きく変わるところがなく、その詳細については説明を省略する。

【００７５】図１６において、各ディスクアレイサブセット１０は、複数の論理ユニット（ＬＵ）１１０で構成されている。各ＬＵ１１０は、独立ＬＵとして構成される。一般に、各ディスクアレイサブセット１０内のＬＵ１１０に割り当てられるＬＵＮは、０から始まる連続番号である。このため、ホスト３０に対して、ディスクアレイシステム1内のすべてのＬＵ１１０のＬＵＮを連続的に見せる場合には、第１実施形態と同様に、フレームペイロード４０２のＬＵＮフィールドを変換する必要がある。本実施形態では、各ディスクアレイサブセット１０のＬＵＮをそのままホスト３０に見せることで、フレームペイロード４０２の変換を不要とし、ディスクアレイスイッチの制御を簡単なものとしている。

【００７６】本実施形態のディスクアレイスイッチ２０は、ホストＩ／Ｆノード２０３ごとに特定のディスクアレイサブセット１０をアクセスできるものと仮定する。この場合、一つのホストＩ／Ｆ３１を使うと、１台のディスクアレイサブセット１０にあるＬＵ１１０のみがアクセス可能である。１台のホストから複数のディスクアレイサブセット１０のＬＵ１１０をアクセスしたい場合には、そのホストを複数のホストＩ／Ｆノード２０３に接続する。また、複数のホスト３０から１台のディスクアレイサブセット１０のＬＵ１１０をアクセスできるようにする場合は、同一のホストＩ／Ｆノード２０３にループトポロジーや、ファブリックトポロジー等を用い、複数のホスト３０を接続する。このように構成すると、１台のホスト３０から１つのＬＵ１１０をアクセスする際に、ホストＩ／Ｆノード２０３のD_ID毎にディスクアレイサブセット１０が確定することになるため、各ＬＵのＬＵＮをそのままホスト３０に見せることが可能である。

【００７７】本実施形態では、上述した理由により、ホスト３０に、各ディスクアレイサブセット１０内のＬＵ１１０のＬＵＮをそのままホスト３０に見せているため、ディスクアレイスイッチ２０におけるＬＵＮの変換は不要となる。このため、ディスクアレイスイッチ２０は、ホスト３０からフレームを受信すると、フレームヘッダ４０１のみを第１実施例と同様にして変換し、フレームペイロード４０２は変換せずにディスクアレイサブセット１０に転送する。本実施形態における各部の動作は、フレームペイロード４０２の変換が行われないことを除くと第１実施形態と同様であるので、ここでは詳細な説明を省略する。本実施形態によれば、ディスクアレイスイッチ２０の開発を容易にできる。

【００７８】［第３実施形態］第２実施形態では、ディスクアレイスイッチのホストＩ／Ｆノードにおいて、フレームヘッダのみを変換しているが、以下に説明する第３実施形態ではフレームヘッダも含め、フレームの変換を行わない形態について説明する。本実施形態のコンピュータシステムは、図１に示す第１実施形態におけるコンピュータシステムと同様に構成される。

【００７９】第１、および第２実施形態では、ホスト３０に対し、ディスクアレイサブセット１０の台数や、ＬＵ１１０の構成等、ディスクアレイシステム1の内部構成を隠蔽している。このため、ホスト３０からはディスクアレイシステム1が全体で１つの記憶装置として見える。これに対し、本実施形態では、ディスクアレイサブセット１０をそのままホスト３０に公開し、ホスト３０がフレームヘッダのD_IDとして直接ディスクアレイサブセットのポートのＩＤを使えるようにする。これにより、ディスクアレイスイッチは、フレームヘッダの情報に従ってフレームの転送を制御するだけで済み、従来技術におけるファイバチャネルのファブリック装置と同等のスイッチ装置をディスクアレイスイッチ２０に替えて利用することができる。

【００８０】ディスクアレイシステム構成管理手段７０は、ディスクアレイサブセット１０の通信コントローラ１０６、及びディスクアレイスイッチ２０の通信手段２０４と通信して各ディスクアレイサブセット１０及びディスクアレイスイッチ２０の構成情報を獲得し、あるいは、設定する。

【００８１】ディスクアレイスイッチ２０は、基本的には図３に示す第１実施形態におけるディスクアレイスイッチと同様の構成を有する。しかし、本実施形態では、ホスト３０が発行するフレームのフレームヘッダの情報をそのまま使ってフレームの転送を制御するため、第１

実施形態、あるいは第２実施形態でディスクアレイスイッチ２０のホストＩ／Ｆノード２０３、ディスクアレイＩ／Ｆノード２０２が有するＤＣＴ２０２７や、ＳＣ２０２２、ＳＰＧ２０２４等により実現されるフレームヘッダ等の変換の機能は不要となる。ディスクアレイスイッチ２０が有するクロスバスイッチ２０１は、フレームヘッダの情報に従ってホストＩ／Ｆノード２０３、及びディスクアレイＩ／Ｆノード２０２の間でファイバチャネルのフレームの転送を行う。

【００８２】本実施形態では、ディスクアレイシステムの構成をディスクアレイシステム構成管理手段７０で一括して管理するために、ディスクアレイ管理用テーブル（以下、このテーブルもＤＣＴと呼ぶ）をディスクアレイシステム構成管理手段７０に備える。ディスクアレイシステム構成管理手段７０が備えるＤＣＴは、図６、７に示す、システム構成テーブル20270とサブセット構成テーブル202720～202723の２つのテーブル群を含む。なお、本実施形態では、ホストＬＵは全てＩＬＵとして構成されるため、ホストＬＵ構成テーブル20271のLU Typeは全て「ＩＬＵ」となり、CLU Class、CLU Stripe Sizeは意味をなさない。

【００８３】管理者は、管理端末５を操作してディスクアレイシステム構成管理手段７０と通信し、ディスクアレイサブセット１０のディスク容量、ディスクユニットの台数等の情報を得て、ディスクアレイサブセット１０のＬＵ１１０の設定、ＲＡＩＤレベルの設定等を行う。次に管理者は、管理端末５によりディスクアレイシステム構成管理手段７０と通信し、ディスクアレイスイッチ２０を制御して、各ホスト３０とディスクアレイサブセット２０間の関係情報を設定する。

【００８４】以上の操作により、ディスクアレイシステム1の構成が確立し、ホスト３０から管理者が望む通りにＬＵ１１０が見えるようになる。ディスクアレイ構成管理手段７０は以上の設定情報を保存し、管理者からの操作に応じ構成の確認や、構成の変更を行うことができる。

【００８５】本実施形態によれば、ひとたびディスクアレイシステム1を構成すれば、管理者からディスクアレイスイッチ２０の存在を認識させることが無く、複数のディスクアレイサブシステムを１台のディスクアレイシステムと同様に扱うことができる。また、本実施形態によれば、ディスクアレイスイッチ２０とディスクアレイサブセット１０は、同一の操作環境によって統一的に操作することができ、その構成確認や、構成変更も容易になる。さらに、本実施形態によれば、従来使用していたディスクアレイシステムを本実施形態におけるディスクアレイシステムに置き換える場合に、ホスト３０の設定を変更することなく、ディスクアレイシステム1の構成をそれまで使用していたディスクアレイシステムの構成に合わせることができ、互換性を維持できる。

【００８６】［第４実施形態］以上説明した第１から第３の実施形態では、ホストＩ／Ｆにファイバチャネルを使用している。以下に説明する実施形態では、ファイバチャネル以外のインタフェースが混在した形態について説明する。

【００８７】図１７は、ホストＩ／ＦがパラレルＳＣＳＩである場合のホストＩ／Ｆノード２０３内部のＩＣ２０２３の一構成例を示す。20230はパラレルＳＣＳＩのプロトコル制御を行うＳＣＳＩプロトコルコントローラ（ＳＰＣ）、20233はファイバチャネルのプロトコル制御を行うファイバチャネルプロトコルコントローラ（ＦＰＣ）、20231はパラレルＳＣＳＩとファイバチャネルのシリアルＳＣＳＩをプロトコル変換するプロトコル変換プロセッサ（ＰＥＰ）、20232はプロトコル変換中データを一時保存するバッファ（ＢＵＦ）である。

【００８８】本実施形態において、ホスト３０は、ディスクアレイＩ／Ｆノード２０３に対してＳＣＳＩコマンドを発行する。リードコマンドの場合、ＳＰＣ20230は、これをＢＵＦ20232に格納し、ＰＥＰ20231に割り込みでコマンドの受信を報告する。ＰＥＰ20231は、ＢＵＦ20232に格納されたコマンドを利用し、ＦＰＣ20233へのコマンドに変換し、ＦＰＣ20233に送る。ＦＰＣ20233は、このコマンドを受信すると、フレーム形式に変換し、ＳＣ２０２２に引き渡す。この際、エクスチェンジＩＤ、シーケンスＩＤ、ソースＩＤ、デスティネイションＩＤは、以降の処理が可能なようにＰＥＰ20231により付加される。あとのコマンド処理は、第１実施形態と同様に行われる。

【００８９】ディスクアレイサブセット１０は、データの準備が完了すると、データ転送準備完了フレームの発行、データ転送、正常終了後ステータスフレームの発行を実施する。ディスクアレイサブセット１０からＩＣ２０２３までの間では、フレームヘッダ４０１やフレームペイロード４０２が必要に応じ変換されながら、各種フレームの転送が行われる。ＩＣ２０２３のＦＰＣ20233は、データ転送準備完了フレームを受信し、続いてデータを受信してＢＵＦ20232に格納し、続けて正常に転送が終わったならば、ステータスフレームを受信し、ＰＴＰ20231に割り込みをかけてデータの転送完了を報告する。ＰＴＰ20231は、割り込みを受けると、ＳＰＣ20230を起動し、ホスト３０に対しデータ転送を開始するよう指示する。ＳＰＣ20230はホスト３０にデータを送信し、正常終了を確認するとＰＴＰ20231に対し割り込みで正常終了を報告する。

【００９０】ここでは、ファイバチャネル以外のホストＩ／Ｆの例としてパラレルＳＣＳＩを示したが、他のインタフェース、例えば、メインフレームへのホストＩ／ＦであるESCON等に対しても同様に適用することが可能である。ディスクアレイスイッチ２０のホストＩ／Ｆノード２０３として、例えば、ファイバチャネル、パラレ

ルＳＣＳＩ、及びESCONに対応したホストＩ／Ｆノードを設けることで、１台のディスクアレイシステム1に、メインフレームと、パーソナルコンピュータ、ワークステーション等のいわゆるオープンシステムの両方を混在させて接続することが可能である。本実施形態では、ディスクアレイＩ／Ｆとしては、第１から第３実施形態と同様、ファイバチャネルを用いているが、ディスクアレイＩ／Ｆに対しても任意のＩ／Ｆを使用することが可能である。

【００９１】［第５実施形態］次に、ディスクアレイシステム1の構成管理の方法について、第５実施形態として説明する。図１８は、本実施形態のシステム構成図である。本実施形態では、ホスト３０が4台設けられている。ホスト“＃０”、“＃１”とディスクアレイシステム1の間のＩ／Ｆ３０はファイバチャネル、ホスト“＃２”とディスクアレイシステム1の間は、パラレルＳＣＳＩ（Ultra SCSI）、ホスト“＃３”とディスクアレイシステム1の間は、パラレルＳＣＳＩ（Ultra2 SCSI）で接続されている。

【００９２】パラレルＳＣＳＩのディスクアレイスイッチ２０への接続は第４実施形態と同様に行われる。ディスクアレイシステム1は、４台のディスクアレイサブセット３０を有する。ディスクアレイサブセット“＃０”には４つの独立ＬＵ、ディスクアレイサブセット“＃１”には２つの独立ＬＵがそれぞれ構成されている。ディスクアレイサブセット“＃２”と“＃３”で１つの統合ＬＵが構成されている。本実施形態では、第１実施形態と同様、ホスト３０に対しディスクアレイサブセット１０を隠蔽し、ファイバチャネルのフレームを変換するものとする。各ＬＵに割り当てられるＬＵＮは、ディスクアレイサブセット“＃０”のＬＵから順に、ＬＵＮ＝０、１、２、・・・６までの７つである。

【００９３】図１９は、管理端末５の表示画面上に表示される画面の一例である。図は、ホストＩ／Ｆ３１と各論理ユニット（ＬＵ）との対応を示した論理接続構成画面である。

【００９４】論理接続構成画面５０には、各ホストＩ／Ｆ３１に関する情報３１００、各ＬＵ１１０に関する情報11000、ディスクアレイサブセット１０とＬＵ１１０の関係等が表示される。ホストＩ／Ｆ３１に関する情報としては、Ｉ／Ｆ種類、Ｉ／Ｆ速度、ステータス等が含まれる。ＬＵ１１０に関する情報としては、格納サブセット番号、ＬＵＮ、容量、ＲＡＩＤレベル、ステータス、情報、等が表示される。管理者はこの画面を参照することで、容易にディスクアレイシステム１の構成を管理することができる。

【００９５】論理接続構成画面５０上で、ホストＩ／ＦとＬＵの間に引かれている線は、各ホストＩ／Ｆ３１を経由してアクセス可能なＬＵ１１０を示している。ホストＩ／Ｆから線の引かれていないＬＵ１１０に対して、そのホストＩ／Ｆに接続するホスト３０からはアクセスできない。ホスト３０によって、扱うデータ形式が異なり、また使用者も異なることから、セキュリティ維持上、適切なアクセス制限を設けることが不可欠である。そこで、システムを設定する管理者が、この画面を用いて、各ＬＵ１１０とホストＩ／Ｆとの間のアクセス許可をあたえるか否かによって、アクセス制限を実施する。図において、例えば、ＬＵ“＃０”は、ホストＩ／Ｆ“＃０”および“＃１”からアクセス可能であるが、ホストＩ／Ｆ“＃２”、“＃３”からはアクセスできない。ＬＵ“＃４”は、ホストＩ／Ｆ“＃２”からのみアクセス可能である。

【００９６】このようなアクセス制限を実現するためアクセス制限情報は、ディスクアレイシステム構成管理手段７０からディスクアレイスイッチ２０に対して送信される。ディスクアレイスイッチ２０に送られたアクセス制限情報は、各ホストＩ／Ｆノード２０３に配信され、各ホストＩ／Ｆノード２０３のＤＣＴ２０２７に登録される。ホストにより、アクセスが制限されたＬＵに対するＬＵ存在有無の検査コマンドが発行された場合、各ホストＩ／Ｆノード２０３は、ＤＣＴ２０２７の検査を行い、検査コマンドに対し応答しないか、あるいは、エラーを返すことで、そのＬＵは、ホストからは認識されなくなる。ＬＵ存在有無の検査コマンドとしては、ＳＣＳＩプロトコルの場合、Test Unit Readyコマンドや、Inquiryコマンドが一般に用いられる。この検査なしに、リード／ライトが実施されることはないため、容易にアクセスの制限をかけることが可能である。

【００９７】本実施形態ではホストＩ／Ｆ３１毎にアクセス制限をかけているが、これを拡張することで、ホスト３０毎にアクセス制限をかけることも容易に実現できる。また、ホストＩ／Ｆ３１、ホスト３０、あるいは、アドレス空間を特定して、リードのみ可、ライトのみ可、リード／ライトとも可、リード／ライトとも不可といった、コマンドの種別に応じたアクセス制限をかけることもできる。この場合、アクセス制限情報としてホストＩ／Ｆ番号、ホストＩＤ、アドレス空間、制限コマンド等を指定してディスクアレイスイッチ２０に制限を設定する。

【００９８】次に、新たなディスクアレイサブセット１０の追加について説明する。ディスクアレイサブセット１０を新規に追加する場合、管理者は、ディスクアレイスイッチ２０の空いているディスクアレイＩ／Ｆノード２０２に追加するディスクアレイサブセット１０を接続する。つづけて、管理者は、管理端末５を操作し、論理接続構成画面５０に表示されている「最新状態を反映」ボタン５００１を押下する。この操作に応答して、未設定のディスクアレイサブセットを表す絵が画面上に表示される（図示せず）。このディスクアレイサブセットの絵が選択されるると、ディスクアレイサブセットの設

定画面が現れる。管理者は、表示された設定画面上で、新規に追加されたディスクアレイサブセットの各種設定を実施する。ここで設定される項目にはＬＵの構成、ＲＡＩＤレベル等がある。続けて、図１９の論理接続構成図の画面に切り替えると、新規ディスクアレイサブセットとＬＵが現れる。以降、ホストＩ／Ｆ３１毎に対するアクセス制限を設定し、「設定実行」ボタン５００２を押下すると、ディスクアレイスイッチ２０に対し、アクセス制限情報、およびディスクアレイサブセット、ＬＵの情報が転送され、設定が実行される。

【００９９】各ディスクアレイサブセット１０にＬＵ１１０を追加する際の手順も上述した手順で行われる。また、ディスクアレイサブセット、およびＬＵの削除についてもほぼ同様の手順で行われる。異なる点は、管理者が各削除部位を画面上で選択して「削除」ボタン５００３を押下し、適切な確認が行われたのち、実行される点である。以上のように、管理端末７０を用いることで、管理者はディスクアレイシステム全体を一元的に管理できる。

【０１００】［第６実施形態］次に、ディスクアレイスイッチ２０によるミラーリングの処理について、第６実施形態として説明する。ここで説明するミラーリングとは、２台のディスクアレイサブセットの２つの独立ＬＵにより二重書きをサポートする方法であり、ディスクアレイサブセットのコントローラまで含めた二重化である。従って、信頼性は、ディスクのみの二重化とは異なる。

【０１０１】本実施形態におけるシステムの構成は図１に示すものと同じである。図１に示す構成おいて、ディスクアレイサブセット“＃０”と“＃１”は全く同一のＬＵ構成を備えており、この２つのディスクアレイサブセットがホスト３０からは１つのディスクアレイとして見えるものとする。便宜上、ミラーリングされたディスクアレイサブセットのペアの番号を“＃０１”と呼ぶ。また、各ディスクアレイサブセットのＬＵ“＃０”とＬＵ“＃１”によってミラーリングペアが形成され、このＬＵのペアを便宜上、ＬＵ“＃０１”と呼ぶ。ＤＣＴ２０２７のホストＬＵ構成テーブル20271上でＬＵ＃０１を管理するための情報は、CLU Classに「Mirrored」が設定され、LU Info.として、ＬＵ＃０とＬＵ＃１に関する情報が設定される。その他の各部の構成は第１実施形態と同様である。

【０１０２】本実施形態における各部の動作は、第１実施例とほぼ同様である。以下、第１実施形態と相違する点について、ディスクアレイスイッチ２０のホストＩ／Ｆノード２０３の動作を中心に説明する。図２０は、本実施形態におけるライト動作時に転送されるフレームのシーケンスを示す模式図、図２１、２２は、ライト動作時におけるホストＩ／Ｆノード２０３による処理の流れを示すフローチャートである。

【０１０３】ライト動作時、ホスト３０が発行したライトコマンドフレーム（FCP_CMD）は、ＩＣ２０２３により受信される（図２０の矢印（ａ）：ステップ21001）。ＩＣ２０２３により受信されたライトコマンドフレームは、第１実施形態で説明したリード動作時におけるステップ20002　20005と同様に処理される（ステップ21002 - 21005）。

【０１０４】ＳＣ２０２２は、ＳＰ２０２１を使ってＤＣＴ２０２７を検索し、ミラー化されたディスクアレイサブセット“＃０１”のＬＵ“＃０１”へのライトアクセス要求であることを認識する（ステップ21006）。ＳＣ２０２２は、ＦＢ２０２５上に、受信したコマンドフレームの複製を作成する（ステップ21007）。ＳＣ２０２２は、ＤＣＴ２０２７に設定されている構成情報に基づいてコマンドフレームの変換を行い、ＬＵ“＃０”とＬＵ“＃１”の両者への別々のコマンドフレームを作成する（ステップ21008）。ここで、ＬＵ“＃０”を主ＬＵ、ＬＵ“＃１”を従ＬＵと呼び、コマンドフレームにもそれぞれ主コマンドフレーム、従コマンドフレームと呼ぶ。そして、両者別々にＥＴ２０２６にエクスチェンジ情報を格納し、ディスクアレイサブセット“＃０”およびディスクアレイサブセット“＃１”に対し作成したコマンドフレームを発行する（図２０の矢印（ｂ０）（ｂ１）：ステップ21009）。

【０１０５】各ディスクアレイサブセット“＃０”、“＃１”は、コマンドフレームを受信し、それぞれ独立にデータ転送準備完了フレーム（FCP_XFER_RDY）をディスクアレイスイッチ２０に送信する（図２０の矢印（ｃ０）（ｃ１））。ディスクアレイスイッチ２０では、ホストＩ／Ｆノード２０３が、第１実施形態におけるリード動作のステップ20011　20013と同様の処理により転送されてきたデータ転送準備完了フレームを処理する（ステップ21011 - 21013）。

【０１０６】各ディスクアレイサブセットからのデータ転送準備完了フレームがそろった段階で（ステップ21014）、ＳＣ２０２２は、主データ転送準備完了フレームに対する変換を実施し（ステップ21015）、ＩＣ２０２３により変換後のフレームをホスト３０に送信する（図２０の矢印（ｄ）：ステップ21015）。

【０１０７】ホスト３０は、データ転送準備完了フレームを受信した後、ライトデータ送信のため、データフレーム（FCP_DATA）をディスクアレイスイッチ２０に送信する（図２０の矢印（ｅ））。ホスト３０からのデータフレームは、ＩＣ２０２３により受信されると（ステップ21031）、リードコマンドフレームやライトコマンドフレームと同様に、ＦＢ２０２５に格納され、ＣＲＣ検査、フレームヘッダの解析が行われる（ステップ21032、21033）。フレームヘッダの解析結果に基づき、ＥＴ２０２６がＳＰ２０２１により検索され、エクスチェンジ情報が獲得される（ステップ21034）。

【0108】SC2022は、ライトコマンドフレームのときと同様に複製を作成し（ステップ21035）、その一方をディスクアレイサブセット“#0”内のLU“#0”に、他方をディスクアレイサブセット“#1”内のLU“#1”に向けて送信する（図20の矢印（f0）（f1）：ステップ21037）。

【0109】ディスクアレイサブセット“#0”、“#1”は、各々、データフレームを受信し、ディスクユニット104に対しそれぞれライトし、ステータスフレーム（FCP_RSP）をディスクアレイスイッチ20に送信する。

【0110】SC2022は、ディスクアレイサブセット“#0”、“#1”それぞれからステータスフレームを受信すると、それらのステータスフレームから拡張ヘッダを外してフレームヘッダを再現し、ET2026からエクスチェンジ情報を獲得する（ステップ21041、21042）。

【0111】ディスクアレイサブセット“#0”、“#1”の両者からのステータスフレームが揃うと（ステップ21043）、ステータスが正常終了であることを確認のうえ、LU“#0”からの主ステータスフレームに対する変換を行い（ステップ21044）、従ステータスフレーム消去する（ステップ21045）。そして、IC2023は、正常終了を報告するためのコマンドフレームをホストに送信する（図20の矢印（h）：ステップ21046）。最後にSP2021は、ET2026のエクスチェンジ情報を消去する（ステップ21047）。

【0112】以上でミラーリング構成におけるライト処理が終了する。ミラーリングされたLU“#01”に対するリード処理は、データの転送方向が異なるだけで、上述したライト処理とほぼ同様に行われるが、ライトとは異なり、2台のディスクアレイサブセットにリードコマンドを発行する必要はなく、どちらか一方に対してコマンドフレームを発行すればよい。たとえば、常に主LUに対してコマンドフレームを発行してもよいが、高速化のため、主／従双方のLUに対して、交互にコマンドフレームを発行するなどにより、負荷を分散すると有効である。

【0113】上述した処理では、ステップ21014、及びステップ21043で2台のディスクアレイサブセット“#0”、“#1”の応答を待ち、両者の同期をとって処理が進められる。このような制御では、双方のディスクアレイサブセットでの処理の成功が確認されてから処理が進むため、エラー発生時の対応が容易になる。その一方で、全体の処理速度が、どちらか遅いほうの応答に依存してしまうため、性能が低下するという欠点がある。

【0114】この問題を解決するため、ディスクアレイスイッチにおいて、ディスクアレイサブセットの応答を待たずに次の処理に進んだり、ディスクアレイサブセットのどちらか一方からの応答があった時点で次の処理に進む「非同期型」の制御をすることも可能である。非同期型の制御を行った場合のフレームシーケンスの一例を、図20において破線矢印で示す。

【0115】破線矢印で示されるフレームシーケンスでは、ステップ21016で行われるホストへのデータ転送準備完了フレームの送信が、ステップ21009の処理の後、ディスクアレイサブセット10からのデータ転送準備完了フレームを待たずに実施される。この場合、ホストに送信されるデータ転送準備完了フレームは、ディスクアレイスイッチ20のSC2022により生成される（破線矢印（d′））。

【0116】ホスト30からは、破線矢印（e′）で示されるタイミングでデータフレームがディスクアレイスイッチ20に転送される。ディスクアレイスイッチ20では、このデータフレームが一旦FB2025に格納される。SC2022は、ディスクアレイサブセット10からのデータ転送準備完了フレームの受信に応答して、データ転送準備完了フレームが送られてきたディスクアレイサブセット10に対し、FB2025に保持されたデータフレームを転送する（破線矢印（f0′）、（f1′））。

【0117】ディスクアレイスイッチ20からホスト30への終了報告は、双方のディスクアレイサブシステム10からの報告（破線矢印（g0′）、（g0′））があった時点でおこなわれる（破線矢印（h′））。このような処理により、図20に示される時間Taの分だけ処理時間を短縮することが可能である。

【0118】ディスクアレイスイッチ20とディスクアレイサブセット10間のフレーム転送の途中でエラーが発生した場合、以下の処理が実施される。

【0119】実行中の処理がライト処理の場合、エラーが発生したLUに対し、リトライ処理が行われる。リトライが成功すれば、処理はそのまま継続される。あらかじめ設定された規定の回数のリトライが失敗した場合、ディスクアレイスイッチ20は、このディスクアレイサブセット10（もしくはLU）に対するアクセスを禁止し、そのことを示す情報をDCT2027に登録する。また、ディスクアレイスイッチ20は、MP200、通信コントローラ204を経由して、ディスクシステム構成手段70にそのことを通知する。

【0120】ディスクシステム構成手段70は、この通知に応答して管理端末5にアラームを発行する。これにより管理者は、トラブルが発生したことを認識できる。その後、ディスクアレイスイッチ20は、正常なディスクアレイサブセットを用いて運転を継続する。ホスト30は、エラーが発生したことを認識することはなく、処理を継続できる。

【0121】本実施形態によれば、2台のディスクアレイサブシステムでミラー構成を実現できるので、ディスクの耐障害性を上げることことができる。また、ディス

クアレイコントローラ、ディスクアレイＩ／Ｆ、及びディスクアレイＩ／Ｆノードの耐障害性を上げることができ、内部バスの二重化等するくとなくディスクアレイシステム全体の信頼性を向上させることができる。

【０１２２】［第７実施形態］次に、３台以上のディスクアレイサブセット１０を統合し、１台の論理的なディスクアレイサブセットのグループを構成する方法について説明する。本実施形態では、複数のディスクアレイサブセット１０にデータを分散して格納する。これにより、ディスクアレイサブセットへのアクセスを分散させ、特定のディスクアレイサブセットへのアクセスの集中を抑止することで、トータルスループットを向上させる。本実施形態では、ディスクアレイスイッチによりこのようなストライピング処理を実施する。

【０１２３】図２３は、本実施形態におけるディスクアレイシステム1のアドレスマップである。ディスクアレイサブセット１０のアドレス空間は、ストレイプサイズＳでストライピングされている。ホストから見たディスクアレイシステム１のアドレス空間は、ストライプサイズＳ毎に、ディスクアレイサブセット“＃０”、“＃１”、“＃２”、“＃３”に分散されている。ストライプサイズＳのサイズは任意であるが、あまり小さくない方がよい。ストライプサイズＳが小さすぎると、アクセスすべきデータが複数のストライプに属するストライプまたぎが発生したときに、その処理にオーバヘッドが発生するおそれがある。ストライプサイズＳを大きくすると、ストライプまたぎが発生する確率が減少するので性能向上のためには好ましい。ＬＵの数は任意に設定することができる。

【０１２４】以下、本実施形態におけるホストＩ／Ｆノード２０３の動作について、図２４に示す動作フローチャートを参照しつつ第１実施形態との相違点に着目して説明する説明する。なお、本実施形態では、ＤＣＴ２０２７のホストＬＵ構成テーブル20271上で、ストライピングされたホストＬＵに関するCLU Classには「Striped」が、CLU Stripe Sizeにはストライプサイズ「Ｓ」が設定される。

【０１２５】ホスト３０がコマンドフレームを発行すると、ディスクアレイスイッチ２０は、ホストＩ／Ｆノード２０３のＩＣ２０２３でこれを受信する（ステップ２２００１）、ＳＣ２０２２は、ＩＣ２０２３からこのコマンドフレームを受け取り、ＳＰ２０２１を使ってＤＣＴ２０２７を検索し、ストライピングする必要があることを認識する（ステップ２２００５）。

【０１２６】次に、ＳＣ２０２２は、ＳＰ２０２１によりＤＣＴ２０２７を検索し、ストライプサイズＳを含む構成情報から、アクセスの対象となるデータが属するストライプのストライプ番号を求め、このストライプがどのディスクアレイサブセット１０に格納されているか特定する（ステップ22006）。この際、ストライプまたぎが発生する可能性があるが、この場合の処理については後述する。ストライプまたぎが発生しない場合、ＳＰ２０２１の計算結果に基づき、ＳＣ２０２２はコマンドフレームに対し変換を施し（ステップ22007）、エクスチェンジ情報をＥＴ２０２６に格納する（ステップ22008）。以降は、第１実施形態と同様の処理が行われる。

【０１２７】ストライプまたぎが発生した場合、ＳＰ２０２１は、２つのコマンドフレームを生成する。この生成は、例えば、ホスト３０が発行したコマンドフレームを複製することで行われる。生成するコマンドフレームのフレームヘッダ、フレームペイロード等は、新規に設定する。第６実施形態と同様、ＳＣ２０２２でコマンドフレームの複製を作成した後、変換を実施することも可能であるが、ここでは、ＳＰ２０２１により新規に作成されるものとする。ＳＣ２０２２は、２つのコマンドフレームが生成されると、これらを各ディスクアレイサブセット１０に送信する。

【０１２８】この後、第１実施形態と同様にデータ転送が実施される。ここで、本実施形態では、第１実施形態、あるいは第６実施形態と異なり、データ自体を１台のホスト３０と２台のディスクアレイサブセット１０間で転送する必要がある。たとえば、リード処理の場合、２台のディスクアレイサブセット１０から転送されるデータフレームは、すべてホスト３０に転送する必要がある。この際ＳＣ２０２２は、各ディスクアレイサブセット１０から転送されてくるデータフレームに対し、ＥＴ２０２６に登録されたエクスチェンジ情報に従い、適切な順番で、適切なエクスチェンジ情報を付加してホスト３０に送信する。

【０１２９】ライト処理の場合は、コマンドフレームの場合と同様、２つのデータフレームに分割して、該当するディスクアレイサブセット１０に転送する。なお、データフレームの順序制御は、ホスト、あるいはディスクアレイサブセットがアウトオブオーダー（Out of Order）機能と呼ばれる、順不同処理に対応しているならば必須ではない。

【０１３０】最後に、すべてのデータ転送が完了し、ディスクアレイスイッチ２０が２つのステータスフレームをディスクアレイサブセット１０から受信すると、ＳＰ２０２１（あるいはＳＣ２０２２）は、ホスト３０へのステータスフレームを作成し、これをＩＣ２０２３によりホスト３０に送信する。

【０１３１】本実施形態によれば、アクセスを複数のディスクアレイサブセットに分散することができるので、トータルとしてスループットを向上させることができるとともに、アクセスレイテンシも平均的に低減させることが可能である。

【０１３２】［第８実施形態］次に、２台のディスクアレイシステム（またはディスクアレイサブセット）間における複製の作成について、第８実施形態として説明す

る。ここで説明するようなシステムは、２台のディスクアレイシステムの一方を遠隔地に配置し、天災等による他方のディスクアレイシステムの障害に対する耐性を備える。このような災害に対する対策をディザスタリカバリと呼び、遠隔地のディスクアレイシステムとの間で行われる複製の作成のことをリモートコピーと呼ぶ。

【０１３３】第６実施形態で説明したミラーリングでは、地理的にほぼ同一の場所に設置されたディスクアレイサブセット１０でミラーを構成するので、ディスクアレイＩ／Ｆ２１はファイバチャネルでよい。しかし、リモートコピーを行うディスクアレイ（ディスクアレイサブセット）が１０ｋｍを越える遠隔地に設置される場合、中継なしでファイバチャネルによりフレームを転送する事ができない。ディザスタリカバリに用いられる場合、お互いの間の距離は通常数百ｋｍ以上となる、このため、ファイバチャネルでディスクアレイ間を接続することは実用上不可能であり、ＡＴＭ（Asynchronous Transfer Mode）等による高速公衆回線や衛星通信等が用いられる。

【０１３４】図２５は、本実施形態におけるディザスタリカバリシステムの構成例である。

【０１３５】８１はサイトＡ、８２はサイトＢであり、両サイトは、地理的な遠隔地に設置される。9は公衆回線であり、ＡＴＭパケットがここを通過する。サイトＡ８１、およびサイトＢ８２は、それぞれディスクアレイシステム1を有する。ここでは、サイトＡ８１が通常使用される常用サイトであり、サイトＢ８２はサイトＡ８１が災害等でダウンしたときに使用されるリモートディザスタリカバリサイトである。

【０１３６】サイトＡ８１のディスクアレイシステム１０のディスクアレイサブセット“＃０”、“＃１”の内容は、サイトＢ８２のディスクアレイシステム１０のリモートコピー用ディスクアレイサブセット“＃０”、“＃１”にコピーされる。ディスクアレイスイッチ２０のＩ／Ｆノードのうち、リモートサイトに接続するものはＡＴＭを用いて公衆回線9に接続されている。このノードをＡＴＭノード２０５と呼ぶ。ＡＴＭノード２０５は、図５に示すホストＩ／Ｆノードと同様に構成され、ＩＣ２０２３がＡＴＭ－ファイバチャネルの変換を行う。この変換は、第４実施形態におけるＳＣＳＩ－ファイバチャネルの変換と同様の方法により実現される。

【０１３７】本実施形態におけるリモートコピーの処理は、第６実施形態におけるミラーリングの処理と類似する。以下、第６実施形態におけるミラーリングの処理と異なる点について説明する。

【０１３８】ホスト３０がライトコマンドフレームを発行すると、サイトＡ８１のディスクアレイシステム１０は、第６実施形態における場合と同様にフレームの二重化を実施し、その一方を自身のディスクアレイサブセット10に転送する。他方のフレームは、ＡＴＭノード２０５によりファイバチャネルフレームからＡＴＭパケットに変換され、公衆回線9を介してサイトＢ８２に送られる。

【０１３９】サイトＢ８２では、ディスクアレイスイッチ２０のＡＴＭノード２０５がこのパケットを受信する。ＡＴＭノード２０５のＩＣ２０２３は、ＡＴＭパケットからファイバチャネルフレームを再現し、ＳＣ２０２２に転送する。ＳＣ２０２２は、ホスト３０からライトコマンドを受信したときと同様にフレーム変換を施し、リモートコピー用のディスクアレイサブセットに転送する。以降、データ転送準備完了フレーム、データフレーム、ステータスフレームのすべてにおいて、ＡＴＭノード２０５においてファイバチャネルーＡＴＭ変換を行い、同様のフレーム転送処理を実施することにより、リモートコピーが実現できる。

【０１４０】ホスト３０がリードコマンドフレームを発行した際には、ディスクアレイスイッチ２０は、自サイトのディスクアレイサブセット１０に対してのみコマンドフレームを転送し、自サイトのディスクアレイサブセット１０からのみデータをリードする。このときの動作は、第１実施形態と同一となる。

【０１４１】本実施形態によれば、ユーザデータをリアルタイムでバックアップし、天災等によるサイト障害、ディスクアレイシステム障害に対する耐性を備えることができる。

【０１４２】［第９実施形態］次に、一台のディスクアレイサブセット１０に包含される複数のＬＵの統合について説明する。例えば、メインフレーム用のディスク装置は、過去のシステムとの互換性を維持するために、論理ボリュームのサイズの最大値が２ＧＢに設定されている。このようなディスクアレイシステムをオープンシステムでも共用する場合、ＬＵは論理ボリュームサイズの制限をそのまま受けることになり、小サイズのＬＵが多数ホストから見えることになる。このような方法では、大容量化が進展した場合に運用が困難になるという問題が生じる。そこで、ディスクアレイスイッチ２０の機能により、この論理ボリューム（すなわちＬＵ）を統合して一つの大きな統合ＬＵを構成することを考える。本実施形態では、統合ＬＵの作成をディスクアレイスイッチ２０で実施する。

【０１４３】本実施形態におけるＬＵの統合は、第１実施形態における複数のディスクアレイサブセット１０による統合ＬＵの作成と同一である。相違点は、同一のディスクアレイサブセット１０内の複数ＬＵによる統合であることだけである。ディスクアレイシステムとしての動作は、第１実施形態と全く同一となる。

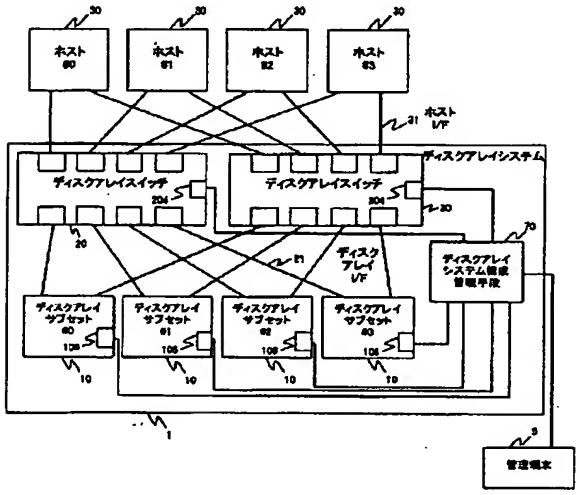
【０１４４】このように、同一のディスクアレイサブセット１０に包含される複数のＬＵを統合して一つの大きなＬＵを作成することで、ホストから多数のＬＵを管理する必要がなくなり、運用性に優れ、管理コストを低減

したディスクアレイシステムを構築できる。
【０１４５】［第１０実施形態］次に、ディスクアレイスイッチ１０による交代パスの設定方法について、図２６を参照しつつ説明する。
【０１４６】図２６に示された計算機システムにおける各部の構成は、第１の実施形態と同様である。ここでは、２台のホスト３０が、各々異なるディスクアレイＩ／Ｆ２１を用いてディスクアレイサブセット１０をアクセスするとように構成していると仮定する。図では、ディスクアレイサブセット、ディスクアレイスイッチ２０のホストＩ／Ｆノード２０３およびディスクアレイＩ／Ｆノード２０２は、ここでの説明に必要な数しか示されていない。
【０１４７】ディスクアレイサブセット１０は、図２と同様の構成を有し、２つのディスクアレイＩ／Ｆコントローラはそれぞれ１台のディスクアレイスイッチ２０に接続している。ディスクアレイスイッチ２０の各ノードのＤＣＴ２２７には、ディスクアレイＩ／Ｆ２１の交代パスが設定される。交代パスとは、ある一つのパスに障害が発生した場合にもアクセス可能になるように設けられる代替のパスのことである。ここでは、ディスクアレイＩ／Ｆ“＃０”の交替パスをディスクアレイＩ／Ｆ“＃１”、ディスクアレイＩ／Ｆ“＃１”の交替パスをディスクアレイＩ／Ｆ“＃０”として設定しておく。同様に、ディスクアレイサブセット１０内の上位アダプタ間、キャッシュ・交代メモリ間、下位アダプタ間のそれぞれについても交代パスを設定しておく。
【０１４８】次に、図２６に示すように、ディスクアレイサブセット１の上位アダプタ“＃１”に接続するディスクアレイＩ／Ｆ２１が断線し、障害が発生したと仮定して、交替パスの設定動作を説明する。このとき、障害が発生したディスクアレイＩ／Ｆ２１を利用しているホスト“＃１”は、ディスクアレイサブセット１０にアクセスできなくなる。ディスクアレイスイッチ２０は、ディスクアレイサブセット１０との間のフレーム転送の異常を検出し、リトライ処理を実施しても回復しない場合、このパスに障害が発生したと認識する。
【０１４９】パスの障害が発生すると、ＳＰ２０２１は、ＤＣＴ２０２７にディスクアレイＩ／Ｆ“＃１”に障害が発生したことを登録し、交代パスとしてディスクアレイＩ／Ｆ“＃０”を使用することを登録する。以降、ホストＩ／Ｆノード２０３のＳＣ２０２２は、ホスト“＃１”からのフレームをディスクアレイＩ／Ｆ“＃０”に接続するディスクアレイＩ／Ｆノード２０２に転送するように動作する。
【０１５０】ディスクアレイサブセット１０の上位アダプタ１０１は、ホスト“＃１”からのコマンドを引き継いで処理する。また、ディスクアレイスイッチ２０は、ディスクアレイシステム構成管理手段７０に障害の発生を通知し、ディスクアレイシステム構成管理手段７０により管理者に障害の発生が通報される。
【０１５１】本実施形態によれば、パスに障害が発生した際の交替パスへの切り替えを、ホスト側に認識させることなく行うことができ、ホスト側の交代処理設定を不要にできる。これにより、システムの可用性を向上させることができる。
【０１５２】以上説明した各実施形態では、記憶メディアとして、すべてディスク装置を用いたディスクアレイシステムについて説明した。しかし、本発明は、これに限定されるものではなく、記憶メディアとしてディスク装置に限らず、光ディスク装置、テープ装置、ＤＶＤ装置、半導体記憶装置等を用いた場合にも同様に適用できる。
【０１５３】
【発明の効果】本発明によれば、計算機システムの規模、要求などに応じた記憶装置システムの拡張、信頼性の向上などを容易に実現することのできる記憶装置システムを実現することができる。
【図面の簡単な説明】
【図１】第１実施形態のコンピュータシステムの構成図である。
【図２】第１実施形態のディスクアレイサブセットの構成図である。
【図３】第１実施形態のディスクアレイスイッチの構成図である。
【図４】第１実施形態におけるディスクアレイスイッチのクロスバスイッチの構成図である。
【図５】第１実施形態におけるディスクアレイスイッチのホストＩ／Ｆノードの構成図である。
【図６】システム構成テーブルの構成図である。
【図７】サブセット構成テーブルの構成図である。
【図８】ファイバチャネルのフレームの構成図である。
【図９】ファイバチャネルのフレームヘッダの構成図である。
【図１０】ファイバチャネルのフレームペイロードの構成図である。
【図１１】ホストからのリード動作時にファイバチャネルを通して転送されるフレームのシーケンスを示す模式図である。
【図１２】ホストＬＵ、各ディスクアレイサブセットのＬＵ、及び各ディスクユニットの対応関係を示す模式図である。
【図１３】ライト処理時のホストＩ／Ｆノードにおける処理のフローチャートである。
【図１４】スイッチングパケットの構成図である。
【図１５】複数のディスクアレイスイッチをクラスタ接続したディスクアレイシステムの構成図である。
【図１６】第２実施形態におけるコンピュータシステムの構成図である。
【図１７】第４実施形態におけるディスクアレイスイッ

チのインタフェースコントローラの構成図である。
【図１８】第５実施形態におけるコンピュータシステムの構成図である。
【図１９】論理接続構成画面の表示例を示す画面構成図である。
【図２０】第６実施形態におけるフレームシーケンスを示す模式図である。
【図２１】第６実施形態のミラーリングライト処理時のホストＩ／Ｆノードにおける処理のフローチャートである。
【図２２】第６実施形態のミラーリングライト処理時のホストＩ／Ｆノードにおける処理のフローチャートである。
【図２３】第７実施形態におけるホストＬＵと各ディスクアレイサブセットのＬＵとの対応関係を示す模式図である。
【図２４】第７実施形態におけるホストＩ／Ｆノードの処理を示すフローチャートである。
【図２５】第８実施形態におけるディザスタリカバリシステムの構成図である。
【図２６】交替パスの設定についての説明図である。
【符号の説明】
1…ディスクアレイシステム、５…管理端末、１０…ディスクアレイサブセット、２０…ディスクアレイスイッチ、３０…ホストコンピュータ、７０…ディスクアレイシステム構成管理手段、２００…管理プロセッサ、２０１…クロスバスイッチ、２０２…ディスクアレイＩ／Ｆノード、２０３…ホストＩ／Ｆノード、２０４…通信コントローラ。

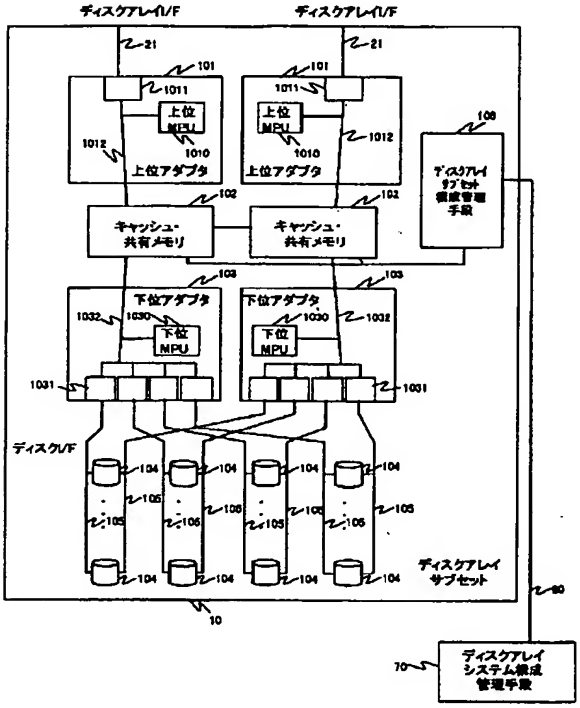
【図１】

図1



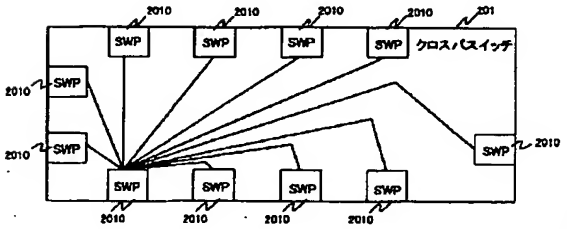
【図２】

図2



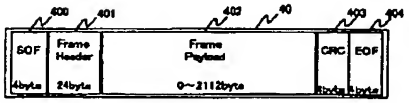
【図４】

図4



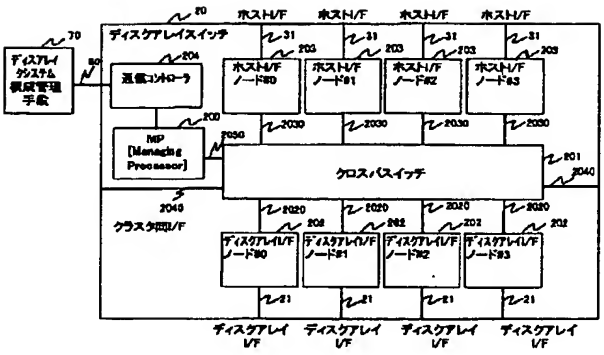
【図８】

図8



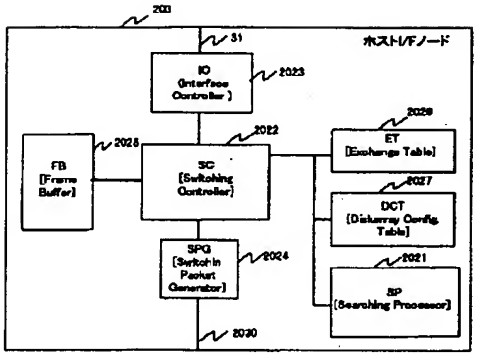
【図3】

図3



【図5】

図5



【図6】

図6

システム構成テーブル

ホストLUテーブル

Host-LU No	LU Type	CLU Class	CLU Stripe Size	Condition	LU Info			LU Info			LU Info			LU Info		
					Subset	LUN	Size	Subset	LUN	Size	Subset	LUN	Size	Subset	LUN	Size
0	CLU	Joined	—	Normal	#0	0	n0	#1	0	n1	#2	0	n2	#3	0	n3
1	—	—	—	Not Defined	—	—		—	—		—	—		—	—	
2	—	—	—	Not Defined	—	—		—	—		—	—		—	—	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

ディスクアレイI/Fノード構成テーブル

Subset	Subset Port No.	Switch No	I/F Node No.
#0	0	0	#0
#0	1	1	#0
#1	0	0	#1
#1	1	1	#1
⋮	⋮	⋮	⋮

【図9】

図9

R_CTL	D_ID
未使用	S_ID
Type	F_CTL
SEQ_ID	DF_CTL / SEQ_CNT
OX_ID	RX_ID
Parameters	

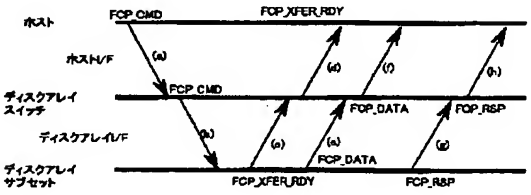
【図10】

図10

LUN (High)
LUN (Low)
CNTL
CDB (word0)
CDB (word1)
CDB (word2)
CDB (word3)
Data Length

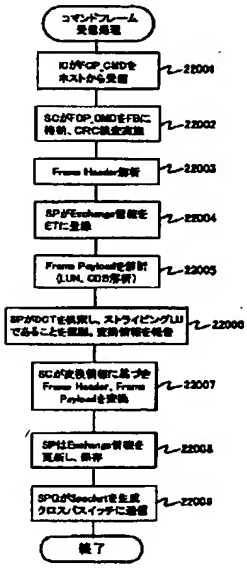
【図11】

図11



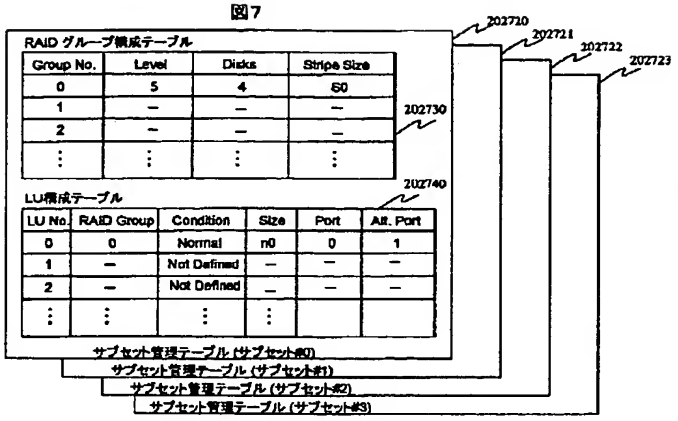
【図24】

図24



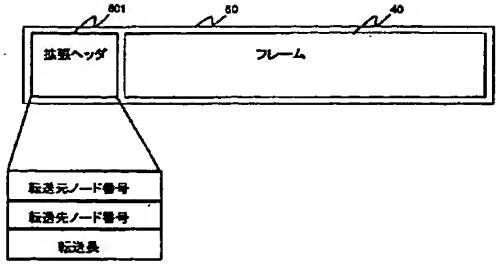
【図7】

図7



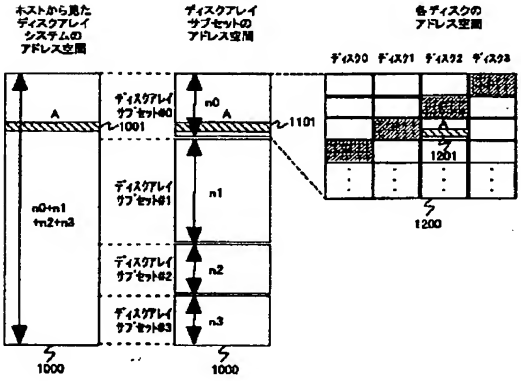
【図14】

図14



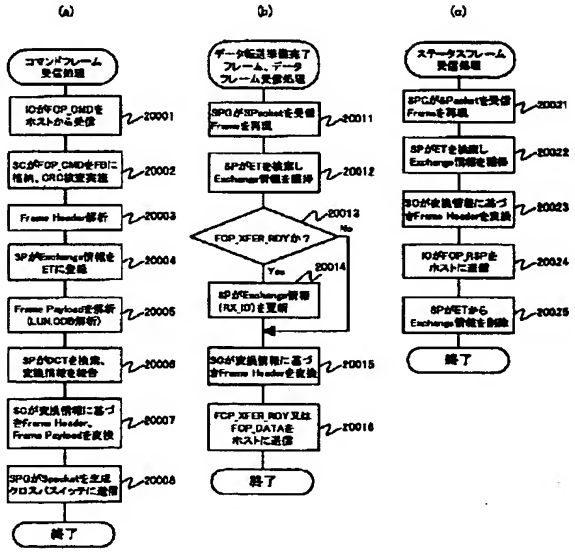
【図12】

図12



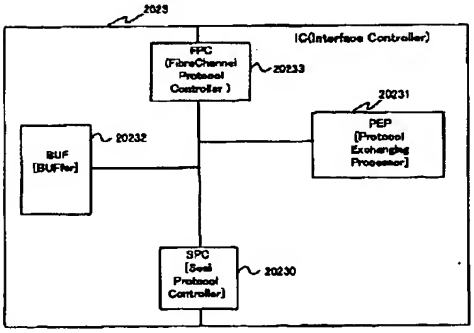
【図13】

図13



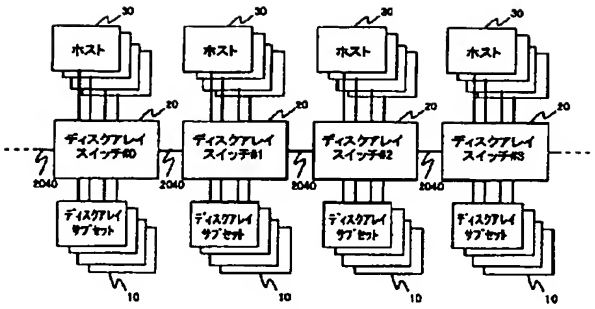
【図17】

図17



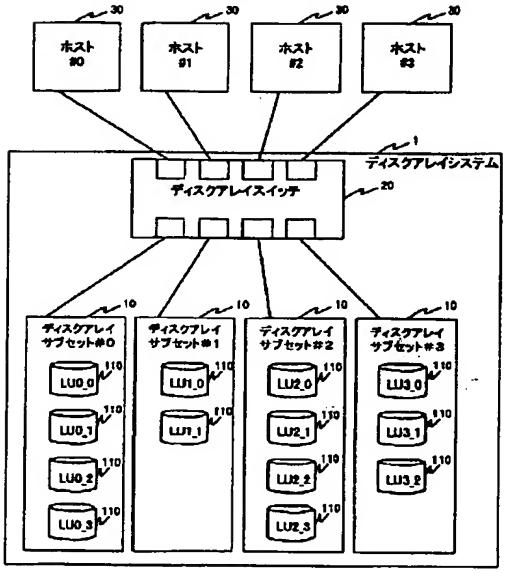
【図15】

図15



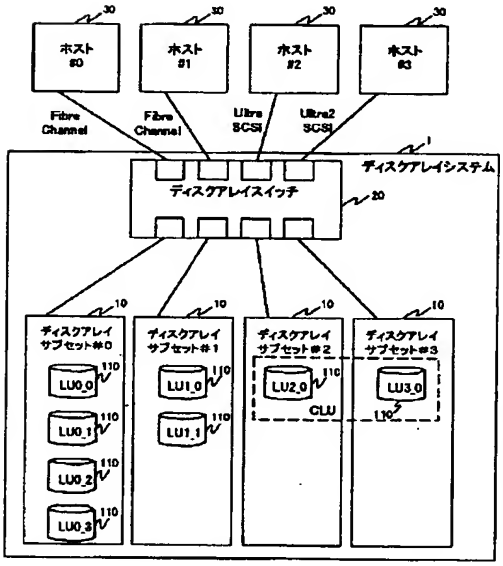
【図16】

図16



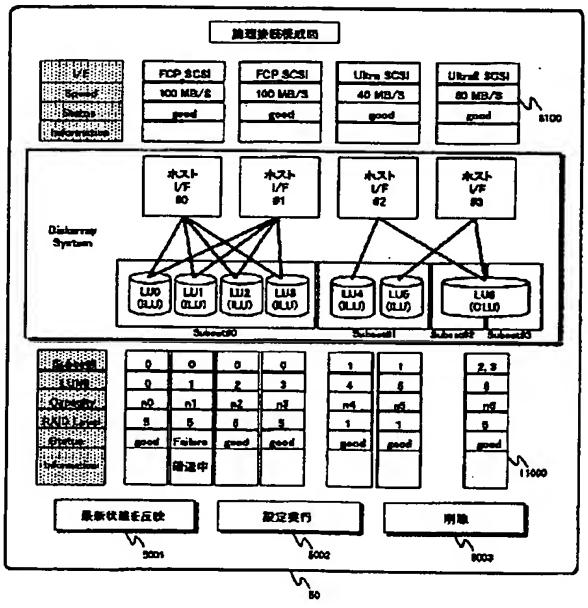
【図18】

図18



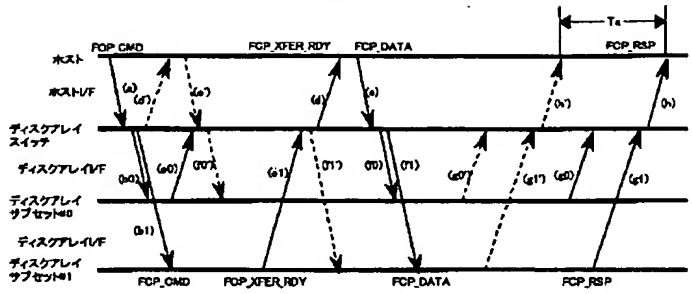
【図19】

図19



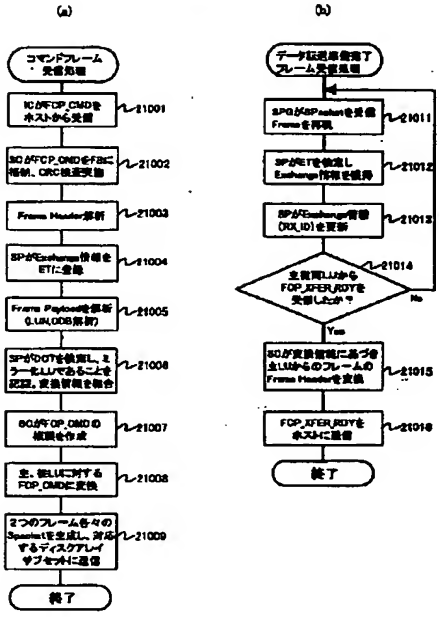
【図20】

図20



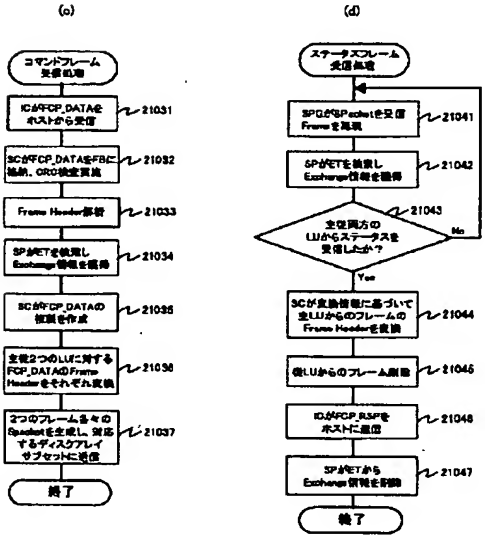
【図21】

図21



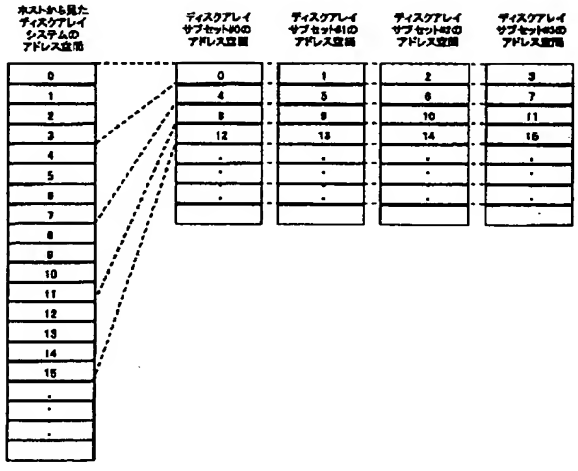
【図22】

図22



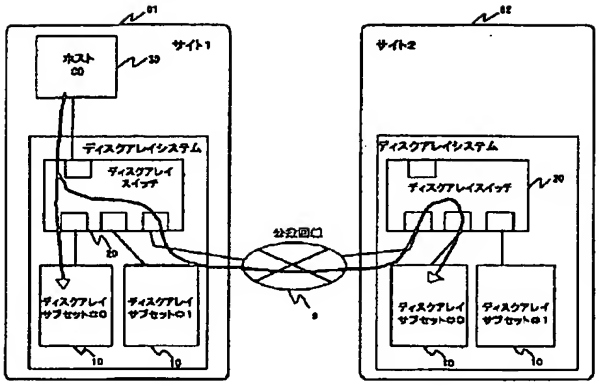
【図23】

図23



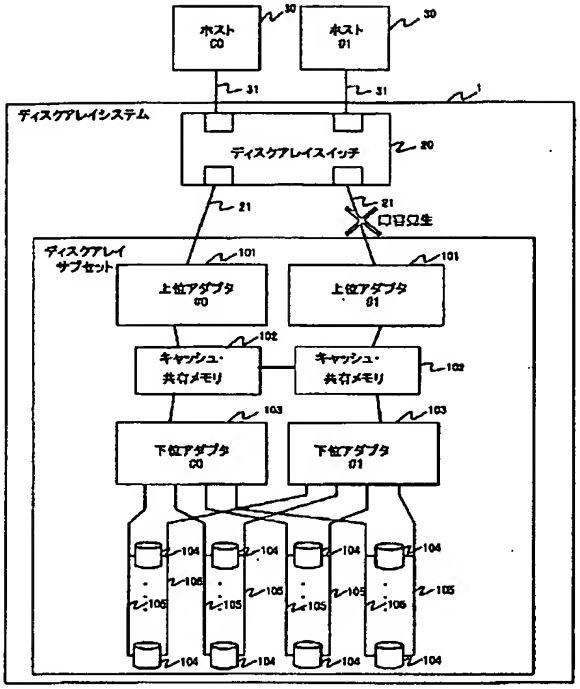
【図25】

図25



【図26】

図26



フロントページの続き

(72)発明者 山本 彰
神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所システム開発研究所内

(72)発明者 味松 康行
神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所システム開発研究所内

(72)発明者 佐藤 雅彦
神奈川県小田原市国府津2880番地 株式会社日立製作所ストレージシステム事業部内